

THE 'WRITE' CUE – INVESTIGATING THE
FACILITATION AND THERAPY EFFECTS OF ORTHOGRAPHIC CUES
IN SPOKEN WORD-FINDING REHABILITATION FOR
ADULTS WITH ACQUIRED APHASIA

WEI PING SZE

B.Soc.Sci. (Hons.) and MSocSci, Psychology, National University of Singapore
MSc (Speech and Language Pathology), National University of Singapore

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES
UNIVERSITY COLLEGE LONDON (UCL)

2024

‘心记不如笔记’ (Chinese proverb)

‘A memory etched into the heart is still not as reliable as a written record.’

‘Littera scripta manet’ (Latin proverb)

‘The written word will remain.’

Declaration of Authorship

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



Signature of author, Wei Ping SZE

Abstract

Word-finding difficulty (anomia) is commonly observed in individuals with aphasia after stroke. In the first study (Chapter 2), a systematic search was conducted on anomia therapy research published between 2008-2018. Intervention details and outcomes from 222 individuals with spoken word-retrieval deficits post-stroke were extracted and analysed using random forest. Therapy components involving orthographic cues emerged to be strong predictors of successful outcomes on *both* treated and untreated items. However, much remains unclear on what types of orthographic cues will achieve optimal outcomes and how durable their effects are.

Two primary research studies were thus conducted. The first is a facilitation case-series ($n = 5$ adults with anomia post-stroke; Chapters 3 and 4) that compared effects of a single application of four facilitation cues, created by crossing two factors – 2 (Size of orthographic unit: ‘Initial Letter’ versus ‘Whole Word’) x 2 (Modalities: ‘Orthography’ versus ‘Orthography and Phonology’). Outcomes of picture-naming were tracked across four time points (two baselines, 15 minutes and one week post-facilitation). Across group and individual-level participant analyses, facilitation effects of ‘orthography only’ cues (‘Initial Letter: Orthography’ and ‘Whole Word: Orthography’ cues) appeared robust.

Finally, a therapy case-series ($n = 5$ adults with anomia post-stroke; Chapters 5 and 6) clarified the benefits of initial letter and whole word ‘orthography-only’ cues, based on 10 sessions of cueing. Support for both initial letter and whole word written prompts were identified across group and individual-level analyses, although ‘Whole Word: Orthography’ cues improved naming in more dimensions (final naming accuracy, response time) at the group-level and in more individuals. Visual short-term memory span measured by the ‘Corsi Block Tapping Test’ appeared important for initial letter orthographic cues to work. Error analyses (Chapter 7) further revealed that participants made significantly fewer ‘no response(s)’ across time, but this might not be therapy-related. Overall, clinicians are encouraged to administer written cues, where applicable.

Impact Statement

Over 350,000 people in U.K. have aphasia.¹ Amongst the deficits of aphasia, word-finding difficulty (anomia) is the most common. Yet, it remains a puzzle how best to treat spoken word-finding impairment. Different language-based therapies have been developed, but the myriad of therapy options means that selecting the optimal treatment is challenging. To complicate matters, some therapies are under-researched. A systematic search using PubMed revealed that none of the published papers exclusively investigated the efficacy of orthography-based approaches, e.g., written cues, between 2008 - May 2022.

To unlock this therapy puzzle, data from 222 adults with anomia post-stroke were compiled from anomia therapy research published between 2008-2018. Random forest analyses revealed components involving orthographic cues were critical variables underpinning successful word-finding therapies. A facilitation and therapy case-series ($n = 5$ adults with anomia respectively) were undertaken to clarify the perimeters of successful orthographic cueing, including the orthographic unit size (initial letter or whole word), whether the effects would be stronger when orthography was combined with phonology to maximise multimodal input, and the duration of effects.

Findings from this thesis potentially have the following impact.

1. On individuals with aphasia.

One key finding is that orthographic cues can improve the naming of treated items. When written cues are consistently implemented in clinical practice, people with anomia could be beneficiaries – Enhanced abilities in spoken word-retrieval may increase their quality of life by reducing communication breakdowns.

2. On practitioners and their service delivery.

Recent video analyses of actual therapies found that clinicians rarely employ written cues², perhaps partly driven by the lack of specific evidence. This thesis addresses the problem in two concrete ways:

- The meta-analysis combined previously published data from 222 participants and showed that orthographic cueing predicted successful outcomes of spoken picture naming following various naming therapies;
- Detailed case-level evidence of participants provided in Chapters 4 and 6 augment the slender evidence base for orthographic cues. Clinicians can now match their

caseloads against the participants' baseline profiles, thus directly supporting clinical decision making.

3. On aphasia research.

- In Chapter 2, heterogenous small-*n* data were collectively analysed by machine learning. This demonstrates how new innovative techniques, like random forest, can be fruitfully combined with established small-*n* design analyses to deepen investigation.
- Chapters 3 to 7 mark the first time that naming performance is comprehensively analysed using initial/final accuracies, response time and error types. The outcomes in these chapters reinforce the need to capture different naming dimensions for a complete picture of therapy progress.

4. A novel product, the 'RITA' framework (Appendix 2.3)³.

This new framework organises essential components of a language therapy (Regimen, Items, Techniques and Technique Application). It is a useful resource for new researchers, students and newly-qualified practitioners, as they navigate the cornucopia of therapy options. 'RITA' can be adapted for the pediatric population, or for other range indicators beyond language.

Finally, there are plans to disseminate these findings through future publications (journals, news bulletins), social media and presentations in conferences and seminars targeting people with aphasia, healthcare providers, and aphasia researchers.

Endnotes:

¹ Stroke Association (n.d.). *Aphasia awareness*. Data accessed on 12 May 2022 from <https://www.stroke.org.uk/what-is-aphasia/aphasia-awareness>

² Brogan, Godecke, and Ciccone (2020).

³ RITA is also available as Appendix 3 in Sze, Hameau, Warren, and Best (2021).

UCL Research Paper Declaration Form

1. For a research manuscript that has already been published:

- a. **Where was the work published?** (e.g. journal name) Aphasiology
- b. **Who published the work?** (e.g. Elsevier/Oxford University Press)
Taylor and Francis
- c. **When was the work published?** 2021
- d. **Was the work subject to academic peer review?** YES ~~NO~~
- e. **Have you retained the copyright for the work?** ~~YES~~ / NO [If no, please seek permission from the relevant publisher and check the box next to the below statement:
☒ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*

2. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):

- W. P. Sze contributed to and was responsible for all aspects of this paper. These included the initial conception and design of the study, implementation of systematic search, co-screening of eligible articles, extraction and analyses of data, as well as the writing up of the initial draft, subsequent editing and revising the final draft.
- S. Hameau assisted with the co-screening of eligible articles.
- J. Warren contributed to the initial conception, design of the study, and the editing of the paper.
- W. Best contributed to the initial conception, design of the study, and the editing of the paper.

3. In which chapter(s) of your thesis can this material be found? Chapter 2

Candidate's e-signature: [Redacted]	Date: 01 September 2022
Supervisor/senior author(s) e-signature: [Redacted]	Date: 10 October 2022
Supervisor/senior author(s) e-signature: [Redacted]	Date: 11 October 2022

Source: <https://www.grad.ucl.ac.uk/essinfo/guidance-on-selfplagiarism/>

Acknowledgements

Scholarly pursuits, like a PhD dissertation, incur debts of gratitude:

To my primary supervisor, Wendy Best, for her steadfast belief and encouragement throughout the journey. Her knowledge and wise counsel are a precious resource I have gratefully benefitted from. Wendy will no doubt remain a source of inspiration.

To my supervisor, Jane Warren, whose ability to see imperfections in the work through her astute and constructive feedback, is crucial in steering my learning and discovery.

And to my third supervisor, Carol Sacchett, who joined the team later but proved so instrumental on elemental parts of the research. Her keen clinical eye is deeply appreciated.

Also –

To Mickey Dean, for lending his occasional counsel and spreading the word of my two primary research studies – so important when recruiting eligible participants.

To my participants and their caregivers, as well as friends who took part in the norming exercises, this thesis will not be possible without your participation.

To my family for their care and tireless support, as well as my UK-based friends (Michael M., Colin J., Stephen T., Michael K., Bob T., and Andrew M.) for embracing and encouraging me.

And to my peers in the 212 Research Lab, for sharing their knowledge, fears and joy – it is incredible how they each juggle parenthood and graduate studies.

Finally, this work will also not be possible without the generous support from the UCL Graduate Research Scholarship and UCL Overseas Research Scholarship. My PhD journey will not be as smooth otherwise.

I would like to end off by acknowledging that Chapter 2 in this thesis is a version of my work that was previously published in the journal, 'Aphasiology'. Written permission has been granted by the journal editor, Professor Chris Code, on May 4, 2022, to reproduce it in this thesis. The full citation of the published version is printed below, and the UCL Research Declaration Form required by the UCL Doctoral School is found on the previous page.

Full citation of the published article:

Sze, W. P., Hameau, S., Warren, J., & Best, W. (2021). Identifying the components of a successful spoken naming therapy: a meta-analysis of word-finding interventions for adults with aphasia. *Aphasiology*, 35(1), 33-72. <https://doi.org/10.1080/02687038.2020.1781419>

Table of Contents

Declaration of Authorship	i
Abstract	ii
Impact Statement	iii
UCL Research Paper Declaration Form	v
Acknowledgements	vi
Table of Contents	viii
List of Figures	xiii
List of Tables	xv
List of Appendices	xvii
List of Abbreviations	xviii
Chapter 1: Introduction – Overview of thesis	
1.1 Background and chapter outlines	1
1.2 Overall hypothesis and motivation of minimalist designs for primary research studies	3
1.3 The spoken picture naming paradigm and associated language models	3
Chapter 2: Identifying the components of a successful spoken naming therapy — A meta-analysis of word-finding interventions for adults with aphasia	
2.1 Introduction	8
2.1.1 Reviews and meta-analyses	9
2.1.2 Therapy components or ‘active ingredients’	14
2.1.3 Random forest and its utility in word-finding therapy research	32
2.1.4 Purpose of this systematic review and meta-analysis	34
2.2 Method	34
2.2.1 Information sources and search	34
2.2.2 Study selection and eligibility criteria	35
2.2.3 Data extraction for meta-analysis	37
2.2.4 Meta-analysis	39
2.3 Results	40
2.3.1 Successful treatment outcome of treated items post-therapy in the short-term phase	41
2.3.2 Successful treatment outcome of treated items post-therapy during the maintenance phase	41
2.3.3 Successful treatment outcome of untreated items post-therapy in the short-term phase	51
2.4 Discussion	51
Addendum to Chapter 2: Short update on spoken word-finding intervention research between June 2018 and March 2022	56

Chapter 3: Comparing the facilitation effects of orthographic cues with or without the phonological components – Group-level participant analyses

3.1 Introduction	61
3.1.1 Three primary research questions and an overall hypothesis	66
3.1.2 A secondary research question and its hypothesis	68
3.2 Method	68
3.2.1 Design	68
3.2.2 Participants	69
3.2.3 Materials	71
3.2.4 Procedure	72
3.2.4.1 Baseline assessments (Weeks 1 and 2)	72
3.2.4.2 Post-facilitation assessments (Weeks 4 and 5)	73
3.2.4.3 Facilitation session (Week 4)	73
3.2.4.4 Picture naming trials during facilitation	75
3.2.4.5 Data extraction and fidelity check	76
3.2.4.6 Group-level analyses	78
3.2.4.6.1 Group-level analyses for primary research questions (1) to (3): Outcomes of cue conditions across and between time points	78
3.2.4.6.2 Group-level correlational analyses for secondary research question (4): General language processes	79
3.2.4.6.3 Group-level correlational analyses for secondary research question (4): Specific language skills	80
3.3 Results	82
3.3.1 Primary research questions (1) to (3): At the participant group level, which unit and mode of cueing will result in significant improvement; also, how durable are the facilitation effects of each combination of cue (unit x modality)?	82
3.3.2 Secondary research question (4): At the participant group level, what is the relationship between baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?	89
3.3.3 Control task	93
3.4 Discussion	93

Chapter 4: Comparing the facilitation effects of orthographic cues with or without the phonological components – Individual-level participant analyses

4.1 Introduction	97
4.1.1 Role of individual-level analyses and the use of a language processing model	97
4.1.2 Four research questions	100
4.2 Method	102
4.2.1 Design, procedure and stimuli	102
4.2.2 Participants and clinical observations	102
4.2.2.1 P1's clinical profile at baseline and probable locus/loci of impairment	102
4.2.2.2 P2's clinical profile at baseline and probable locus/loci of impairment	108
4.2.2.3 P3's clinical profile at baseline and probable locus/loci of impairment	109
4.2.2.4 P4's clinical profile at baseline and probable locus/loci of impairment	110
4.2.2.5 P5's clinical profile at baseline and probable locus/loci of impairment	112
4.2.3 Individual-level analyses for primary research questions (1) to (3): Outcomes of cue conditions across and between time points	114

4.2.4 Individual-level exploratory analyses for secondary research question (4): Associations between performances on baseline assessments and facilitation outcomes	116
4.3 Results	117
4.3.1 Primary research questions (1) to (3): At the individual participant level, which unit and mode of cueing will result in significant improvement; also, how durable are the facilitation effects of each combination of cue (unit x modality) for each participant?	117
4.3.1.1 'Initial Letter: Orthography' facilitation cue	117
4.3.1.2 'Whole Word: Orthography' facilitation cue	123
4.3.1.3 'Initial Letter: Orthography and Phonology' facilitation cue	123
4.3.1.4 'Whole Word: Orthography and Phonology' facilitation cue	123
4.3.1.5 Distinguishing between improvement based on a facilitation cue versus spontaneous improvement based on repeated exposure	124
4.3.2 Secondary research question (4): What is the relationship between each participant's baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?	125
4.4 Discussion	130

Chapter 5: Comparing the use of initial letter against whole word as the unit of orthographic cue in spoken naming therapy – Group-level analyses

5.1 Introduction	136
5.2 Method	141
5.2.1 Design	141
5.2.2 Participants	143
5.2.3 Procedure	143
5.2.3.1 Assessments (Baseline and post-therapy sessions)	144
5.2.3.2 Therapy sessions	146
5.2.3.3 Picture-naming trials during therapy	147
5.2.3.4 Extraction of main naming outcome measures and fidelity check	148
5.2.3.5 Rating outcomes for participants and caregivers' therapy-related views	149
5.2.3.6 Ethics clearance and impact of Covid-19	149
5.2.4 Group-level analyses	150
5.3 Results	151
5.3.1 Primary research questions (1) and (2): At the participant group level, which unit of the orthographic cue will result in significant improvement; also, how durable are the therapy effects of each orthographic cue?	152
5.3.1.1 Primary analyses on the main naming outcome measures	152
5.3.1.2 Primary supporting analyses on rating outcome measures (Questionnaire responses)	158
5.3.2 Secondary research question (3): At the participant group level, what is the relationship between baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?	160
5.3.3 Control task: Nonword repetition	160
5.4 Discussion	162

Chapter 6: Comparing the use of initial letter against whole word as the unit of orthographic cue in spoken naming therapy – Individual participant analyses

6.1 Introduction	165

6.2 Method	167
6.2.1 Design, procedure and stimuli	167
6.2.2 Participants and clinical observations	167
6.2.2.1 <i>P1's clinical profile at baseline and probable locus/loci of impairment</i>	173
6.2.2.2 <i>P2's clinical profile at baseline and probable locus/loci of impairment</i>	173
6.2.2.3 <i>P3's clinical profile at baseline and probable locus/loci of impairment</i>	175
6.2.2.4 <i>P4's clinical profile at baseline and probable locus/loci of impairment</i>	175
6.2.2.5 <i>P5's clinical profile at baseline and probable locus/loci of impairment</i>	176
6.2.3 Individual analyses for primary research questions (1) and (2): Outcomes of cue conditions across and between time points	177
6.2.4 Individual-level exploratory analyses for secondary research question (3): Associations between performances on baseline assessments and therapy outcomes	178
6.3 Results	179
6.3.1 Primary research questions (1) and (2): At the individual participant level, which unit of the orthographic cue will result in significant improvement; also, how durable are the therapy effects of each orthographic cue?	179
6.3.1.1 <i>Initial letter orthographic cue therapy</i>	179
6.3.2.2 <i>Whole word orthographic cue therapy</i>	187
6.3.2.3 <i>Comparing against improved outcomes from repeated exposure</i>	187
6.3.2.4 <i>Unseen controls</i>	188
6.3.2.5 <i>Control task: Nonword repetition</i>	189
6.3.2 Secondary research question (3): What is the relationship between each participant's baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?	189
6.4 Discussion	191

Chapter 7: Evaluating the changes in error types for initial letter and whole word orthographic cueing – Individual participant error analyses

7.1 Introduction	199
7.2 Method	206
7.2.1 Participants and data collection	206
7.2.2 Classifying naming errors using an error coding scheme	207
7.2.3 Fidelity check	208
7.2.4 Analysing with the non-parametric weighted trend test	210
7.3 Results	211
7.3.1 Target-related naming errors	211
7.3.2 Target-unrelated naming errors	211
7.3.3 Null-type naming errors	218
7.4 Discussion	219
7.4.1 Pooling the error analysis results together	220
7.4.2 Locus of effect for 'repeated exposure'	221
7.4.3 The non-significant effects on target-related and unrelated errors following orthographic cueing	222

Chapter 8: General Discussion

8.1 Summary of main findings	225
8.2 Main contribution (1): New evidence for whole word orthographic cueing – Main findings and putative mechanisms	226
8.3 Main contribution (2): A novel taxonomy to describe therapy components – RITA	229

8.4 Main contribution (3): Applying new analytical techniques to aphasia research, while valuing data from small-n research designs, for fresh perspectives	229
8.5 Main contribution (4): Highlighting the need for naming performance to be measured comprehensively	230
8.6 Discussion issue (1): Language pathway(s) involved in orthographic cueing – The ‘lexical-semantic account (Lorenz & Nickels, 2007) versus sub-lexical account (Best et al., 2002)’ debate	232
8.7 Discussion issue (2): Does orthographic cueing produce generalised improvements to untreated items?	233
8.8 Discussion issue (3): Evidence on the feasibility of online anomia therapy	234
8.9 Limitations and further future directions	235
8.10 Clinical implications	236
References	240
Appendices	267

List of Figures

Figure 1.1a and 1.1b. The two-stage discrete and the interactive cognitive neuropsychological models of single-word naming redrawn from Ellis (1985) and Dell, Schwartz, Martin, Saffran, and Gagnon (1997) respectively	4
Figure 1.2. A more developed cognitive neuropsychological model of single-word naming (adapted from Howard and Harding, 1998; Kay, Coltheart, and Lesser, 1992), with a focus on orthographic and auditory inputs (written and spoken words), ending with spoken output	6
Figure 2.1. PRISMA flow chart for systematic review (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009)	36
Figure 2.2. Variables and their values of variable importance for improved spoken lexical-retrieval of treated items post-therapy in the short-term phase	48
Figure 2.3. Variables and their values of variable importance for improved spoken lexical-retrieval of treated items post-therapy during the maintenance phase	49
Figure 2.4. Variables and their values of variable importance for improved spoken lexical-retrieval of untreated items post-therapy in the short-term phase	50
Figure 2.5. PRISMA flow chart for the update on systematic review searched using PubMed (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009)	57
Figure 3.1. Flowchart of the facilitation study	72
Figure 3.2. Format of assessment and facilitation trials	76
Figure 3.3. Descriptive statistics of group-level results ($n = 5$) for the six conditions across four sessions on each of the three naming outcome measures	84
Figure 4.1. A cognitive neuropsychological model of single-word naming (adapted from Howard and Harding, 1998; Kay, Coltheart, and Lesser, 1992), with a focus on orthographic and auditory language inputs (written and spoken words), ending with spoken output	99
Figure 4.2. Initial naming accuracy scored for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively	118
Figure 4.3. Final naming accuracy scored for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively	119
Figure 4.4. Average RT of correct responses for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively	120
Figure 4.5. Participants' z-scores on baseline language assessments and their groupings according to the facilitation cue(s) found effective. Data sorted based on z-scores	128
Figure 5.1. Flowchart of the therapy study	144
Figure 5.2. Format of assessment and therapy trials	148
Figure 5.3. Descriptive statistics of group-level results ($n = 5$) for each of the conditions and the three main naming outcome measures at the final baseline and	

post-therapy sessions (one week and four weeks post-therapy). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated	153
Figure 5.4. Descriptive statistics of ratings (scale of '1' to '6') on items 1 and 2 from participants and caregivers	159
Figure 6.1. Orthographic input processing (visual word processing) pathways with spoken output	166
Figure 6.2. Initial naming accuracy scored for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated	180
Figure 6.3. Final naming accuracy scored for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated	181
Figure 6.4. Average RT of correct responses for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated	182
Figure 6.5. Participants' z-scores on baseline language and cognitive assessments and their groupings according to the therapy cue(s) found effective. Data sorted based on z-scores	190

List of Tables

Table 1. Brief description of the chapters	1
Table 2.1. Selected reviews and meta-analyses (2009 to 2018)	10
Table 2.2. Information or variables describing the regimen	16
Table 2.3. Information or variables describing the items	17
Table 2.4. Information or variables describing the techniques	19
Table 2.5. Information or variables describing the application of technique(s)	27
Table 2.6. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on treated items, when assessed at the short-term phase after therapy	42
Table 2.7. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on treated items, when assessed during the maintenance phase after therapy	44
Table 2.8. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on untreated items, when assessed at the short-term phase after therapy	46
Table 2.9. Summary of spoken word-finding therapy studies involving semantic, phonological and orthographic language-based approaches published between June 2018 to mid-March 2022	58
Table 3.1. Summary of facilitation studies involving orthographic stimuli only and with spoken output. All studies focused on spoken picture naming accuracy as the primary outcome	62
Table 3.2. Participant information	70
Table 3.3. Results of the group analysis by-participants based on the mean proportional accuracy of their initial naming responses	85
Table 3.4. Results of the group analysis by-participants based on the mean proportional accuracy of their final naming responses	86
Table 3.5. Results of the group analysis by-participants based on the average naming response times	87
Table 3.6. Summary of findings for the cue and repeated exposure conditions for group analyses for each outcome measure	88
Table 3.7. Kendall rank-order correlations between scores based on general language processes or specific baseline language assessments and the proportional naming gains yielded at different post-facilitation time points for each cue condition	90
Table 4.1. Scores on the background language assessments	103
Table 4.2. Summary of participants' language pathways relied on to process written/spoken input, based on their baseline performances	106
Table 4.3. Summary of facilitation findings for each condition for group and individual-level (WEST and nonparametric post-hoc) analyses for each main outcome measure	121

Table 4.4. Changes in the raw number of items correctly named and improvement in average RT (from baseline 2), as well as their proportional changes. For P2, only post hoc results that remained significant after comparison against the corresponding repeated exposure data are reported here	126
Table 4.5. Summary of facilitation cue(s) found effective for each participant and the language pathways they relied on to process written/spoken input, based on their baseline performances	129
Table 5.1. Summary of spoken word-finding therapy studies involving orthographic stimuli published to date. All studies focused on spoken picture naming accuracy as the primary outcome	137
Table 5.2. Participant information	142
Table 5.3. Results of the group analysis by-participants based on the mean proportional accuracy of their initial naming responses	154
Table 5.4. Results of the group analysis by-participants based on the mean proportional accuracy of their final naming responses	155
Table 5.5. Results of the group analysis by-participants based on the average naming response times	156
Table 5.6. Summary of group-level findings on the effectiveness of each condition on the three main naming outcome measures	157
Table 5.7. Kendall rank-order correlations between performance on baseline cognitive assessments and the proportional naming gains yielded across different post-therapy time points for each cue condition	161
Table 6.1. Participants' scores on background language assessments	168
Table 6.2. Participants' scores on the background cognitive assessments	172
Table 6.3. Summary of therapy findings for each condition for group and individual-level (WEST and nonparametric post-hoc) analyses for each main outcome measure. Group-level results are a reproduction of Table 5.6	183
Table 6.4. Changes in the raw number of items correctly named and improvement in average RT (from baseline immediately prior to therapy), and their proportional changes specified within parentheses. Only post hoc results that remained significant after comparison against repeated exposure data are reported	186
Table 7.1. Summary of spoken word-finding facilitation/therapy studies that included error analyses when evaluating a therapy's efficacy. All studies focus on spoken picture naming accuracy as the primary outcome and involve only individuals with aphasia post-stroke	200
Table 7.2. The three categories and their specific types of object-naming errors in aphasia, adapted from Best, Nickels, and Williamson (2005)	209
Table 7.3. Summary of Poisson main and post hoc analyses of each target-related naming error type made by each participant across various time points (pts) for initial naming responses	212
Table 7.4. Summary of Poisson main and post hoc analyses of each target-unrelated and null-type naming error made by each participant across various time points (pts) for initial naming responses	215

List of Appendices

Appendix 2.1. List of information sources	267
Appendix 2.2. Articles eligible for the meta-analysis	268
Appendix 2.3. List of therapy components based on the 'RITA' framework	274
Appendix 3.1. Recruitment flyer used for the facilitation study	275
Appendix 3.2. Brief description of baseline tasks administered in the facilitation study	276
Appendix 3.3. Sources of photographs used as stimuli ($n = 296$) in the facilitation study (Chapters 3 and 4)	278
Appendix 3.4. Changes in the average number of items in raw count correctly named and improvement in average RT (from baseline immediately prior to therapy) (SD) across participants in the facilitation study. Only post hoc results that remained significant after comparison against repeated exposure data are highlighted	279
Appendix 3.5. Table summarizing the post-facilitation changes in terms of values for the studies involving the use of orthographic form as facilitation. Studies highlighted in yellow refer to those whose cue condition closely resembles a facilitation cue condition in Chapters 3 and 4	280
Appendix 4.1. Individual results based on WEST-Trend and ROC for each main naming outcome measure	283
Appendix 4.2. Results for each participant and condition based on item-level WEST analysis on main naming outcomes (i.e., Picture naming accuracies (initial and final responses) and latencies)	293
Appendix 5.1. Brief description of baseline tasks administered during the orthographic cueing therapy study	297
Appendix 5.2a. Post-therapy questionnaire: for the participant with aphasia	301
Appendix 5.2b. Post-therapy questionnaire: for the caregiver	306
Appendix 5.3. Changes in performance on all three main outcome measures using mean proportion (SD) across participants and treated items, as well as changes in performance on accuracy measures using average raw scores (SD) across participants	309
Appendix 6.1. Individual results based on WEST-Trend and ROC for each main naming outcome measure	311
Appendix 6.2. Results for each participant and condition based on item-level WEST analysis on main naming outcomes (i.e., Picture naming accuracies (initial and final responses) and latencies)	321
Appendix 6.3. Participants' z-scores on language and cognitive baseline assessments and their groupings according to the therapy cue(s) found effective. Data sorted based on groupings	331
Appendix 6.4. Table summarizing the post-therapy changes (Δ) in terms of values for the studies involving the use of orthographic cueing. All post-therapy changes (Δ) are significant, when compared against baseline, according to the original papers unless reported otherwise	332
Appendix 7. Raw numbers (maximum number of 'opportunities' for error at each time point) and Poisson main and post hoc analyses of each error type made by each participant across time for initial naming responses	335

List of Acronyms

ANOVA	Analysis of variance
AQ	Aphasia Quotient
BL	Baseline
BNT	Boston Naming Test
CART	Classification and Regression Trees
CAT	Comprehensive Aphasia Test
CV	Consonant-vowel
CVA	Cerebral vascular accident
FN Acc.	Final Naming Accuracy
IN Acc.	Initial Naming Accuracy
MCA	Middle cerebral artery
ms	millisecond
OOB	Out Of Bag
PALPA	Psycholinguistic Assessment of Language Processing in Aphasia
PNT	Philadelphia Naming Test
POLAR	Predicting Outcome of Language Rehabilitation in Aphasia
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomised controlled trial
Resp(s)	Response(s)
RITA	Regimen, Item(s), Technique(s) and Application of technique(s) with their Adjuncts
ROC	Rate of change
RT	Response time
SD	Standard deviation
SFA	Semantic Feature Analysis
SLTs	Speech and language therapists
TEA	Test of Everyday Attention
TIDieR	Template for Intervention Description and Replication
WAB	Western Aphasia Battery
WEST	WEighted STatistics
WMS	Wechsler Memory Scale

Chapter 1

Introduction – Overview of Thesis

1.1 Background and chapter outlines

The genesis of this thesis is rooted in a simple question: “What makes for an optimal spoken word-finding therapy?” Word-finding difficulty, also known as anomia or lexical-retrieval deficit, refers to the failure to retrieve the desired word, resulting in a communication breakdown (Grossman, 2014). It is a core characteristic of aphasia (Helm-Estabrooks & Albert, 2004, p. 43), and remains a residual deficit of many well-recovered individuals with aphasia (Murdoch, 1990, p. 84). Spoken word-finding difficulties can be characterised by semantic, phonological or neologistic paraphasias, use of non-specific words (e.g., “the thing”), circuitous language, or aborted sentences (Tesak & Code, 2008; Martin, 2013). Estimates vary, but based on a comprehensive dataset from July 2013 to July 2015, compiled by a UK national audit register of stroke care – Out of the 88,974 adults who were hospitalised for stroke and survived (data from Scotland excluded), 41% of them ($n = 37,537$) were identified with having aphasia (Mitchell, Gittins, Tyson, Vail, Conroy, Paley, & Bowen, 2021, p. 952). This high prevalence combined with aphasia’s debilitating effects on quality of life (Lam & Wodchis, 2010), mean that the search for practical solutions to optimise aphasia therapy is meaningful and clinically crucial. Since within aphasia therapy, remediating impaired spoken word retrieval is an important mainstay, the speech and language therapist (SLT) is interested in selecting the most effective language therapy components to achieve this outcome.

What emerged from this attempt to answer the initial question, was a detailed investigation of whether orthographic forms should be engaged during word-finding treatment, and for whom orthographic cues would be useful. The work is represented by three interlinked studies, presented here as Chapters 2 to 6. Each study is designed as a response to questions posed by the preceding ones. The chapters are outlined in Table 1.

Table 1. Brief description of the chapters

Chapter 1
The current chapter provides an overview of the thesis, while serving as an introduction/preface.

(continued)

Chapter 2

- This contains a systematic search of over a decade of word-finding rehabilitation research, covering the period from 2008 to 2018. It is followed by a meta-analysis of individual cases ($n = 222$) from published spoken word retrieval studies, which was designed to identify key therapy components underlying successful spoken word-finding treatments.
- Results from this meta-analysis informed the choice of cues for further investigation in the intervention studies reported in subsequent chapters.

Chapters 3 and 4

- These two chapters present a facilitation study, the first of the two primary research studies in this thesis using a case-series approach ($n = 5$ in each study). In the meta-analysis, variables involving the use of orthography were consistently found to be amongst the top components indicative of significant naming improvement (e.g., provision of [part-word] orthographic cue(s), provision of the target word in written form). This directly contradicts the current clinical practice, whereby speech and language therapists are encouraged to administer multimodal cues if possible.
- The purpose of the facilitation study was therefore to test the above meta-analytic finding empirically by evaluating 'orthography only' cues against their multimodal 'orthography and phonology' counterparts.

Chapters 5 and 6

- These two chapters present a therapy study, whose design was informed by the results from the facilitation study. The facilitation demonstrated that outcomes from 'orthography only' cues were robust, but the parameters for successful orthographic cueing remained obscure. These included the optimal unit of cueing and duration of their effects.
- A full intervention study was thus run, contrasting the performance of two written cues from the extreme ends of the orthographic unit spectrum (initial letter versus whole word) at different time intervals.

Chapter 7

- Error analyses were conducted on the data reported in chapter 6. This could shed further light on the impact of the two orthographic cues from a different angle.
-

1.2 Overall hypothesis and motivation of minimalist designs for primary research studies

Based on the meta-analytic finding in Chapter 2, the overarching hypothesis for Chapters 3 to 7, is that orthographic cues would significantly improve participants' naming outcomes. The experimental designs for the facilitation and therapy studies are deliberately kept minimalist, as interventions with fewer components allow clearer conclusions on active ingredients to be drawn (e.g., Howard, Patterson, Franklin, Orchard-Lisle, and Morton, 1985). Another motivation for keeping the intervention pared down is due to ecological validity. In healthcare settings where many SLTs are employed, logistics are often set to a minimal for practical reasons related to decluttering and safety, especially at bedside. The simple, direct orthographic cue used in this work thus aligns neatly with what could work well within the clinical context. The simplicity of cueing is also likely why the process has been well-employed by SLTs for a relatively long time (Linebaugh & Lehner, 1977; Patterson, 2001; Raymer & Roitsch, 2023). It is surprising then that evidence on orthographic cues is particularly meagre, owing to a lack of research. The investigations in the following chapters mark a step towards addressing this surprising dearth of evidence.

1.3 The spoken picture naming paradigm and associated language models

Spoken picture naming is the main paradigm used across all the chapters. In this task, the visual picture will be cognitively translated into a mental referent, which in turn will be transformed into a phonological form for eventual spoken naming. Several cognitive models have been developed to explain this conversion process from mental referent to a spoken word; and some of the more prominent ones are briefly reviewed here. The reason(s) for eventually choosing one of the model(s) for use will also be provided. Interested readers should also consult Rapp and Goldrick (2006) or Biedermann, Fiedel, and Nickels (2018) for comprehensive reviews of related spoken word production models.

A good starting point will be the two-stage feedforward models of single-word naming (i.e., Figure 1.1a, which is redrawn from Ellis, 1985, p. 111; also see Levelt, 1989 and Levelt, Roelofs, & Mayer, 1999, which are mentioned later). The components they contain still underlie much of present-day research. These models (Ellis, 1985; Levelt, 1989) are made up of three primary components, creating two discrete sequential stages and thus their 'two-stage' label. Stage one is lexical selection; firstly, upon viewing a picture, a conceptual representation (referent) is triggered within the conceptual semantic system. Next within stage one, the word representation within the mental dictionary (i.e., speech output lexicon)

that best matches the conceptual representation is identified. In some models, like Levelt (1989), the word representations here are conceptualised as lemmas, which are non-phonological word entries diacritically encoded with semantic and language-specific grammatical information. Examples of the latter include the gender markers found in Dutch and French nouns. In other models, like Ellis (1985), the lexicon is simply a storage of word representations, with each word in a speaker's vocabulary indicated by a single word representation, with no further mention of diacritic encodings (see p. 110 in Ellis, 1985). In the case of spoken production of English nouns, which is the eventual focus of this thesis, additional diacritic encodings are arguably not as applicable, which is consistent with Ellis's model. Stage two is the phonological access stage, whereby specific phonetic forms are attached to that word representation to enable saying aloud. However, one major drawback of these two-stage feedforward models is that they are limited. In the context of this thesis, the discrete two-stage models fail to specify crucial input processes relevant for investigating pathways associated with language cueing, rendering them less useful.

Figure 1.1a and 1.1b. The two-stage discrete and the interactive cognitive neuropsychological models of single-word naming redrawn from Ellis (1985) and Dell, Schwartz, Martin, Saffran, and Gagnon (1997) respectively.

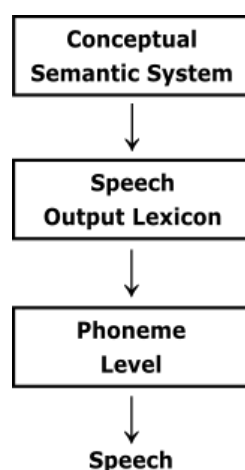


Figure 1.1a

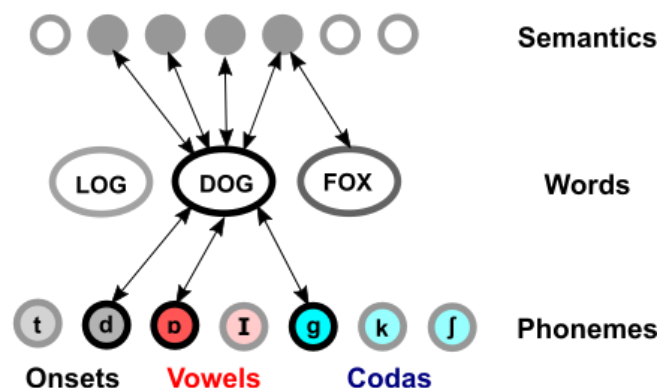


Figure 1.1b

The interactive models by Dell and colleagues (Figure 1.1b; e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000; see Dell, Nozari, & Oppenheim, 2015, pp. 12-14, for a succinct summary) also seek to explain processes underlying spoken picture naming. The same two stages of lexical selection and phonological access are preserved. The former takes place in the semantic and the lemma layers, and the latter in the lemma and phonological layers. A key difference however, from the two-stage discrete models

mentioned above, is that the links between layers in Dell and colleagues' interactive models are *bi-directional*. This means that both feedforward and feedback activations across layers are possible. According to Foygel and Dell (2000), a computerised version of the interactive model parametrised by 's' (semantic) and 'p' (phonological) weights is optimal in terms of modelling the pattern of naming errors produced by individuals with aphasia during spoken picture naming. 's' refers to the strength of connections between the word and semantic nodes; whilst 'p' indicates the strength of connections between the word and phonological nodes. Although adjusting 's' and 'p' weights could in theory, indirectly simulate inputs from semantic and phonological cues, the interactive models still do not account for influences from orthographic inputs – a primary focus of this thesis.

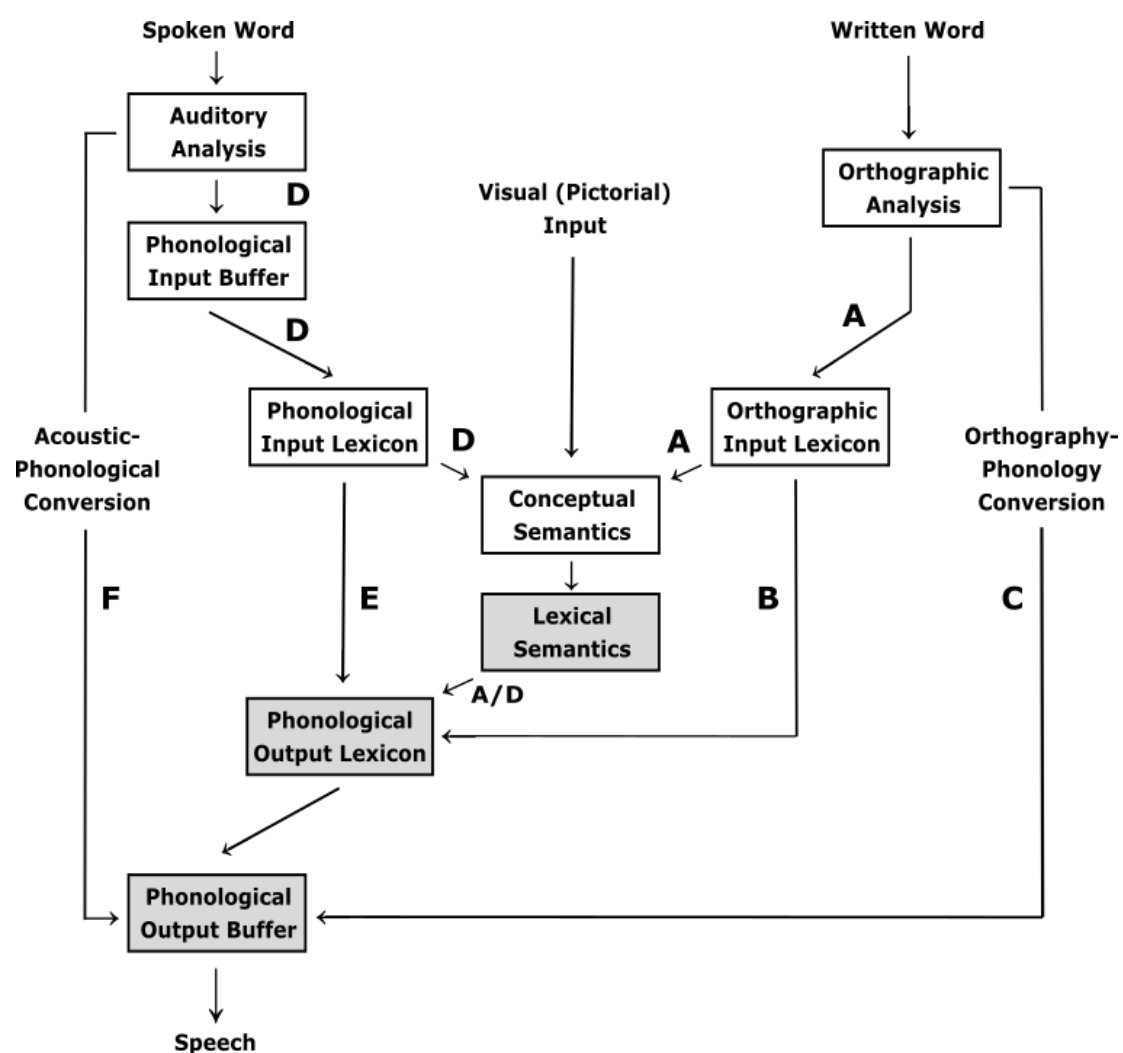
There have been further theoretical advancements made towards refining the two-stage discrete and interactive models. Efforts were made towards detailing the components within the models. These include specifying how the speech output lexicon may be segmented into separate sections for proper nouns (e.g., proper names, like 'Richard'), common nouns (noun stem plus its inflection, e.g., 'book' plus the plural form for 'books'), verbs (verb stem plus its inflection, e.g., 'read' plus '-ing' inflection), and how potentially adding a 'grammatical feature' component could accommodate potential grammatical processes in lexical selection (see p. 57 in Rapp and Goldrick, 2006). Models like WEAVER++ (Word Encoding by Activation and VERification; Roelofs, 2004) further expands on the phoneme level in Figure 1.1b by separating the level into three more levels, namely morpheme nodes, segment nodes (which will be the individual phonemes) and the syllable program nodes (where the phonemes are syllabified into the syllable frames for articulation). More recently, Roelofs (2021) also ran simulations to provide tentative proof of concept that the model could account for phonological cueing in word finding. He sought to demonstrate how the average naming response time generated by WEAVER++ after adjustment, could be similar to the average response time found in Nardo and colleagues' (2017) adults with aphasia (five to 264 months post-stroke) before and after phonological cueing treatment. Nonetheless, the previous criticism of how these models fail to accommodate orthographic inputs, an important consideration for this thesis, still stands.

Finally, most suitable for this thesis will be the cognitive neuropsychology models, like the model underlying the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; Kay, Coltheart and Lesser, 1992). An adapted version is represented below in Figure 1.2.

Figure 1.2 is excerpted from Kay and colleagues (1992) as well as Howard and Harding (1998). Because both source models originate from the same logogen models

designed for single word general processing (e.g., Morton, 1979), they share the same underlying architecture and properties, including the localist framework, discrete components and pathways. Both are also strictly feedforward and threshold models. However, they may label the same components differently. For instance, Kay and colleagues used ‘abstract letter identification’, while Howard and Harding named the same box ‘orthographic analysis’. In Figure 1.2, the labels that directly denote the terms ‘orthography’ and ‘phonology’ are chosen because the terms map directly onto those used later in this thesis, like the facilitation cues of ‘Orthography (only)’ and ‘Multimodal: Orthography and phonology’ in Chapters 3 and 4. Another major difference will be that Kay and colleagues did not distinguish between conceptual and lexical semantics. The distinction is preserved in Figure 1.2, following Howard and Harding (1998).

Figure 1.2. A cognitive neuropsychological model of single-word naming (adapted from Howard and Harding, 1998; Kay, Coltheart, and Lesser, 1992), with a focus on orthographic and auditory inputs (written and spoken words), ending with spoken output.



Crucially, aside from the emphasis on spoken production, the cognitive neuropsychological model also incorporates written and spoken inputs leading to a spoken output. These additional components (i.e., elements on pathways 'A' to 'F' in Figure 1.2; see Chapter 4, section 4.1, for an elaboration on these elements) are particularly pertinent when either written and/or spoken cues are involved during therapy to improve impaired spoken naming. Another reason is because these models are widely applied clinically: Cognitive neuropsychological models are sufficiently detailed and functional, thus commonly used by clinicians in UK, Australia and Singapore, seeking to localise their English-speaking clients' single word language deficits. The assessment outcomes will in turn guide the direction of the clinicians' therapies (Whitworth, Webster, & Howard, 2005).

Core elements from the two-stage models (Figures 1.1a and 1.1b) are still located within Figure 1.2. The two stages of lexical selection and phonological access are processed within three components, lexical semantics, phonological output lexicon and phonological output buffer (see boxes in grey), which are arguably analogous to the components in Figure 1.1a (Ellis, 1985). In lexical semantics, the lexical meanings and associated with the referent will be activated (e.g., 'bark', 'tail', 'kennel' for the referent 'dog'). The phonological output lexicon contains the phonological representations of the words, whilst the phonological output buffer is a temporary storage of the phonological units (e.g., individual phonemes) retrieved and assembled, before the entire word is articulated. In the example of the referent 'dog', during lexical selection, the lexical entry most associated with the activated lexical meanings, i.e., /dɒg/, will be retrieved from the phonological output lexicon. The three constituent phonemes in /dɒg/ are then correspondingly arranged within the output buffer to be spoken aloud. Impairment in any of these components and associated pathways will result in spoken naming difficulty. It appears that Figure 1.2. does not make specific reference on how grammatical information is activated, but this should not negate the model's application since this thesis does not focus on grammatical processing.

Given the relevance of language input cues and need to clinically account for participants' deficits within this thesis, the model depicted in Figure 1.2 (simply referred to as 'the cognitive neuropsychological model' subsequently) is therefore selected. This will help organise case-level analyses in Chapters 4 and 6, by grouping participants according to their common language deficits or residual strengths post-stroke operationalised in terms of the input pathways (i.e., cueing pathways), and how these are related to the particular variations of orthographic input cues that are beneficial (or not).

There is an old Chinese proverb, "the palest ink is superior to the strongest memory." From the perspective of rehabilitation, perhaps the inked text could be an ideal means to re-establish a weakened memory access after a neurological assault. This thesis is an endeavor to find out.

Chapter 2

Identifying the components of a successful spoken naming therapy

— A meta-analysis of word-finding interventions for adults with aphasia

2.1 Introduction

The literature on word-finding treatments is prodigious (Laine & Martin, 2006, pp. 111-131). A PubMed search revealed that between 1973 and 1992, 70 papers were published on the subject matter. The numbers have since risen dramatically: In a similar stretch of twenty years, between 1999 to 2018, there were at least 271 new papers released – almost quadrupling the earlier quantity¹. Understandably, for clinicians with busy schedules, this new abundance overwhelms, rather than enlightens. For each proposed treatment backed by research evidence, there are modifications of the same treatment with similarly compelling empirical provenance. For example, Boyle and Coelho's (1995) original Semantic Feature Analysis (SFA) was designed to treat semantic-level deficits, and in their original research, the participant had to generate six semantic features; in Hashimoto's (2012) modified SFA, only three semantic features were required (also see Boyle, 2010, on further variations, including replacing feature generation with the 'Yes/No' semantic verification tasks). To complicate matters, entirely different approaches to treat the same level of deficit, have been proposed, e.g., the use of written word-to-picture matching task to successfully treat A.E.R.'s semantic deficit (Nickels & Best, 1996b). With evidence-based practice to the fore, it is timely to ask: Which piece of evidence-based therapy method(s) should the SLTs base their practices on? Specifically, what combination of therapy components is most effective?

In this chapter, a meta-analysis of recent literature (2008 - 2018) is conducted, focusing on language-based approaches to treat spoken word retrieval deficits. The structure of this chapter is as follow. First, previous reviews of word-finding therapy are summarised and evaluated. The evaluation clarifies the reasons why this meta-analysis is necessary. Next, therapy components or 'active ingredients' are defined and elaborated. Therapy components refer to specific components that make up a full-fledged language-based therapy. These encompass the more general variables (e.g., related to regimen, like the number of sessions, whether homework should be assigned) and the more treatment-specific variables, like the types of linguistic cues to provide (e.g., semantic or orthographic cues) to achieve a certain rehabilitation goal (e.g., treatment-specific effects in the short

¹ The PubMed search was performed by the author on 08 March 2019.

term). For coherent organisation of the multiple therapy components, the 'RITA' framework is introduced, which stands for (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts. The use of random forest (Breiman, 2001) is then explained, before it is used to systematically navigate through this large-scale dataset ($n = 222$ participants) with the 'RITA' therapy component framework as a structure. Random forest is a well-established method (elaborated below), but whose potential is yet to be fully realised in aphasiology.

The scope of this chapter covers therapies that target spoken single-word production using language forms (i.e., semantics, phonology and orthography), designed for adults with word-finding difficulties after a stroke. Non-stroke populations and interventions espousing the use of alternative communication modes (e.g., alternative communication devices, Beukelman and Mirenda, 2013), total communication strategies (e.g., gestures, Rose & Douglas, 2001; drawings, Hatfield & Zangwill, 1974) or phrase and sentence productions (e.g., exchange of common phrases in constraint-induced language therapy (CILT) by Pulvermüller et al., 2001) are excluded.

2.1.1 Reviews and meta-analyses

Several reviews and meta-analyses of spoken word retrieval therapies for post-stroke adults have been published. In tandem with Chapter 2's objective to summarise literature within the past decade, the following discussion will only include the reviews and meta-analyses published from 2008 onwards. Explanations on why this review is still necessary will also be provided later. Code and Tesak (2008), Nickels and Best (1996a), Nickels (2002b) provide qualitative overviews of earlier research. A quick summary of some recent review papers within our purview, with their more pertinent findings are presented in Table 2.1.²

Each of these reviews attempted to address issues related to effective word-finding therapies, but from different angles and using different methodologies. A shift towards more robust objective syntheses can be discerned, with Casarin and colleagues (2014) being the only review relying on descriptive statistics for analysis. While collectively they add to the richness in word-finding literature, these recent reviews still do not address the issue of the relative importance of active components making up the therapies. Therapy components

² It is beyond the scope of this paper to include all reviews published between 2008 and 2018. This is not intended to be a systematic report of existing reviews of single-word naming therapies.

Table 2.1. Selected reviews and meta-analyses (2009 to 2018).

Paper / Brief description, including aim(s)	Main finding(s)
(1) Wisenburn and Mahoney (2009):	
• 44 articles, based on single-case experimental designs and group studies, published from 1989 and 2007, yielded 107 effect sizes. • The aim was to compare the effect sizes (to represent treatment efficacy) for each of the main word-finding approaches (semantic, phonologic, mixed). Effect sizes were also evaluated based on moderators, like the types of words examined (treated, untreated etc.), and duration (immediately post-therapy, three-months post-therapy etc.).	Subgroup comparisons of effect sizes were computed. • Importantly, the phonological approach (a loose category that includes phonemic cues, orthographic cues, verbal repetition and writing words etc.) yielded the largest magnitude, whether it was immediately post-therapy or in the three-month follow-up. • Also, 'semantic therapy appeared to generalise slightly more than either phonological or mixed therapies to the unexposed words' (p. 1345 - 1346). The authors were however cautious, stating 'no definite conclusions could be made as to whether semantic, phonologic, or mixed therapy is most efficacious' (p. 1346). This is in part because these mean effect sizes combined 'data across all moderator variables' (p. 1344; e.g., participant severity, age, different components across therapies etc.), thus lacking precision to permit firm conclusions.
(2) Snell, Sage, and Lambon Ralph (2010):	
• 21 articles, based on single-case experimental designs, published between 1985 and 2006, yielded 109 individual data points. • The aim was to investigate (a) whether the number of items used in treatment affected the outcome, and (b) whether the severity of the patients' word-finding deficit influenced the outcome.	Correlation analyses were run between naming outcome (computed as 'proportion gain', p. 1067) and (a) number of items used in treatment, as well as (b) severity (operationalised as scores obtained in Boston Naming Test). • The results were not conclusive. This is partly because a bias in literature was identified, i.e., more severe participants were assigned more treatment items. • A follow-up case series based on 13 stroke participants suggested that participants, regardless of severity, can tolerate more items (e.g., $n = 60$) in naming therapies.
(3) Jacquemot, Dupoux, Robotham, and Bachoud-Levi (2012):	
• 39 articles, based on single-case experimental designs, published between 1990 to November 2009, yielded 124 individual data points.	Signal detection theory was used for analysis. The authors argued that sensitivity index (d') was a bias-free measure of rehabilitation sensitivity, and thus appropriate for their

-
- The aim was to identify which processing components (boxes and arrows) within a neuropsychological model are most important for treatment, in order to generate successful outcomes for spoken word-finding. This is different from the other meta-analyses, which examined treatment components, not the 'boxes and arrows' of a theoretical model.

analysis. d' was calculated for each processing component in the neuropsychological model. In this context, d' accounts for whether the treatment outcome is successful and whether the treatment task activates the intended component (e.g., Hits = Successful outcome, task correctly activates intended component; see Jacquemot et al., p. 87).

- Rehabilitation was most successful if the components (a) 'phonological output lexicon' and (b) 'phonological output' were activated.
- Interestingly, if pathways linking (a) 'orthographic input' and 'phonological output', as well as (b) 'orthographic input lexicon' and 'phonological output lexicon' were triggered, rehabilitation was successful too.

However, not all tasks were equally beneficial for treating spoken word-finding. Tasks activating the phonological input and phonological input lexicon (e.g., syllable counting), were not beneficial.

(4) Casarin, Branco, Pereira, Kochhann, Gindri, and Fonseca (2014):

- 28 articles, including those using single-case experimental designs and RCTs, published between 2005 and May 2014 were summarised.
- The aim was to evaluate lexical-semantic interventions (this was defined broadly and included interventions like gestural facilitation training and CILTs)
- However, the target population were not simply adults with stroke, but also those with traumatic brain injury, and/or dementia.
- The outcomes of these studies also do not appear to be specific to spoken single-word naming either: E.g., improvements in functional communication measured by the Amsterdam Nijmegen Everyday Language Test (ANELT) following a broad package of cognitive-linguistic treatment (de Jong-Hagelstein et al., 2011).

Descriptive statistics was used for analyses.

- The authors reported that majority of the studies (39.28% of them) investigated using semantic cues, followed by SFA (17.86%). After these semantic-based interventions, phonological interventions appeared to be next in terms of popularity (17.85%) and then a miscellany of other types of interventions (e.g., intention and pantomime gestures, Copy and Recall Treatment).

(5) Brady, Kelly, Godwin, Enderby and Campbell (2016):

- 57 RCTs, involving 3002 participants, for all studies published until September 2015.
 - 29 subgroup analyses of effect sizes on the outcome of 'expressive language: naming'
-

This, however, includes *any speech and language therapies* seeking to improve communication and participation; and these are *not specific to word-finding*.

- The aim was to provide a comprehensive Cochrane review into speech and language therapy after a stroke.
- The meta-analysis of RCTs examined many topical issues (e.g., therapy versus no therapy) on various outcomes (e.g., functional communication, receptive language, expressive language, economic outcomes etc.). For our purposes, only the subset examining 'expressive language: naming' is of interest.

was done (see Brady et al., pp. 'i' to 'vi', for the complete listing).

- There was no evidence suggesting benefit or harm for most of the comparisons. A selection of these comparisons with 'no difference found' is provided as follow:
 - (a) 'Therapy received' versus 'no therapy received' (this comparison was based on 275 participants across 7 trials; though quality of evidence was rated as 'poor', p. 43);
 - (b) 'High-intensity' versus 'low-intensity' therapy (this was based on 2 studies; intensity was defined as number of hours per week)³;
 - (c) Therapies with short versus long durations (this was based on 3 studies; duration was defined as the total number of weeks)⁴; and
 - (d) Semantically-based therapy versus CILT⁵ (this was just based on 1 study, i.e., Wilssens et al., 2015).
- Interestingly, participants whose therapies were mediated by a computer performed better on measures of untreated items, compared to those in direct contact with an actual therapist (based on 2 studies; p. 38 in Brady et al., 2016).

(6) de Aguiar, Bastiaanse, and Miceli (2016):

- 30 articles, based on single-case experimental designs, published between 1992 to 2014, yielded 166 individual data points.
- The aim was to examine the predictive values of over 30 demographic (e.g., age, gender), clinical (e.g., type of impairment: semantic, sublexical etc.) and treatment-related (semantic cues, phonemic cues, cue direction etc.) variables when they are entered collectively, for two separate outcomes specific to verbs (i.e., improvement on treated vs untreated verbs).

Analysis was completed using random forest.

- For improvement on treated verbs, good verb comprehension, good repetition skills, and number of therapy sessions attended per week were found to be important predictors.
- For improvement on untreated verbs, incorporation of morphological cues in therapy, presence of grammatical impairment, presence of semantic impairment (indexed by noun comprehension) and number of therapy sessions attended per week were found to be important predictors.

³ Intensity was measured in terms of hours per week, and exact operationalization varied between studies (p. 30; Brady et al., 2016). In Brady et al., high-intensity was represented by Pulvermuller (2001) and SP-i-RiT with 10 hours per week and low-intensity being 2 hours in SP-i-RiT and 5 hours in Pulvermuller (2001).

⁴ Duration was measured in terms total number of weeks. Exact operationalization again varied between studies (p. 34; Brady et al., 2016). Short duration in Brady et al. could be two weeks (Pulvermuller, 2001) or between 6 and 9 months (Di Carlo, 1980) and long duration could be between 3 and 5 weeks (Pulvermuller, 2001) or between 5 and 22 months (Di Carlo, 1980).

⁵ Unfortunately, this is the only comparison made between approaches in this review that is pertinent for us (pp. 38-39), and technically, CILT should be excluded from our discussion, as the linguistic unit during therapy is beyond the single-word.

are the nuts and bolts of a treatment protocol. It is these ingredients that speech and language therapists are most interested in, when designing a therapy to treat a client. With the exception of de Aguiar and colleagues (2016), most reviews tended to either frame their discussions of therapies using broad terminologies like phonological or semantic approaches without examining the specific components; or they may simply narrow their focus and manipulate just one component (e.g., number of treatment items by Snell and colleagues, 2010), but in doing so, forced to commit the non-realistic assumption of *ceteris paribus*. Consequently, outstanding issues remain unanswered.

One potential reason why several of these reviews fail to adequately address the makeup of an effective therapy could be methodological. In conventional meta-analyses, synthesis is mathematically accomplished by computing weighted effect sizes for each study, before aggregating them to derive a composite effect (Hartung, Knapp, & Sinha, 2008). In order to examine how one factor (e.g., a particular therapy component) may influence this composite effect, it is common (especially, for meta-analyses in aphasia) to run subgroup analyses as a means to isolate the effects. However, the optimal subgroup analyses (in fact, the optimal meta-analyses) take place only when the studies under inspection are relatively homogenous (see Lipsey & Wilson, 2001, pp. 158-163).

Furthermore, the subgroup analyses tended to be orthogonal (i.e., variables are examined one at a time), but therapies have multiple components with a mix of categorical and continuous units of measurements. Discretising continuous measurements of independent variables will likely incur a loss of information (see Humphreys, 1978, for an analogous examination in the context of ANOVA). Importantly, an analytic approach that focuses on a component, whilst assuming the rest of them are random noise, only masks but does not address possible underlying structural differences embedded across different word-finding intervention studies. This is acknowledged by Wisenburn and Mahoney (2009), who stated that the summary statistic combines effects from both the manipulated variable *and all moderator variables*. 'These moderator variables (in addition to other variables not analysed, such as participant severity, aphasia type, age, [other therapy elements] etc.) may act as a confounding variable in the interpretation of this overall score. It is functionally impossible to eliminate all potential confounding variables from an analysis of aphasia therapy' (p. 1344). Identifying the 'active' therapy ingredients based on these meta-analytic findings confounded by participant and therapy heterogeneity, would be difficult, if not impossible.

In comparison, the approach of de Aguiar and colleagues (2016) is distinctive. It applied a robust technique (random forest) that is novel for the field, and included clinically-

relevant variables for simultaneous evaluation. For all its strengths, there are two points to be made. First, despite the decision to include multiple therapy components as part of their host of variables, the search for active ingredients was not the primary focus of the authors. They cast their net wide and aimed for broad coverage: Demographic variables, various clinical diagnoses, and therapy components are all represented. Including clinical diagnoses might have rendered their analysis murkier. This is because the assessment scores used to inform the clinical diagnoses (e.g., pre-therapy status of sublexical impairment) were derived from diverse tests. The authors are frank about this: ‘Assessment scores ... were obtained from a variety of tests, and even when tests were similar in nature they may have differed in degree of difficulty. This may have introduced noise in our dataset. Unfortunately, many studies report on results of *ad-hoc* tests, but fail to provide normative values for them (p. 12).’ To eschew associated pitfalls including data noise, this analysis will therefore exclude the input of clinical diagnoses as predictors.

Secondly, Auguiar and colleagues did not distinguish between outcomes obtained shortly post-therapy and those obtained during the longer-term maintenance phase. In their paper, these two outcomes were conflated. There are good grounds to keep the two separated. Cognitive processes involved in short term gains, may differ from those responsible for longer-term maintenance (see Cowan, 2008, and Norris, 2017, for recent reviews). Critically, within the aphasia literature, it has been demonstrated that factors (and thus, therapy components) leading to better performance immediately post-therapy, may be distinct from those resulting in superior performance when assessed much later (e.g., Howard et al., 1985). In the present chapter, the intent was to include four outcome measures: i.e., naming performance for treated and untreated items respectively at the short-term and maintenance phases.

It is clear from this overview of recent reviews that outstanding gaps and imperfections remain. These provide motivations for a new systematic review and meta-analysis. Perhaps, the most glaring weakness is the absence of a comprehensive evaluation on the ‘active’ therapy ingredients that underpin an intervention.

2.1.2 Therapy components or ‘active ingredients’

Naming therapies often involve multiple components. These ‘active’ components specify the ‘how’, ‘what’ and to some extent, ‘when’ and ‘where’ (presumably, if therapy is computerised, it could be completed from home; and ‘homework’ as its name implies, is assumed to be done outside therapy time, and likely at home). To this author’s knowledge,

there is no published taxonomy outlining what these therapy components are. Existing checklists, like the 'Template for Intervention Description and Replication' (TIDieR; Hoffmann et al., 2014), provide general guidelines on the items to report. They do not specify the individual components within an intervention.

In Tables 2.2 to 2.5, possible therapy components are listed and organised into four categories: Variables or information about the (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts ('RITA' for an acronym). This 'RITA' framework aims to be comprehensive, not exhaustive, but it should be useful as a spring board for further refinement. Where possible, examples from existing literature (with priority to interventions published within the last decade) are cited in the last column of these tables. Although this framework of therapy components is currently applied to single-word lexical retrieval therapies, the taxonomy can be used for language therapies beyond single-word naming.

The first set provides information on the treatment regimen (Table 2.2). 'Regimen' variables describe the treatment schedule. They are the standard scheduling details clinicians convey to their clients about the expected length of the intervention (i.e., sessional information). Examples are the total number of sessions and the duration of each session etc (also see Warren, Fey, & Yoder, 2007). There is empirical evidence that these variables do influence treatment efficacy. Bhogal, Teasell, and Speechley (2003) combined results from eight studies, and found that increasing the number of hours per week (indexed by number of minutes per session and number of sessions per week in Table 2.2) was significantly related to improved outcomes measured on the Porch Index of Communicative Abilities. To contrast, Auguiar and colleagues (2016) reported that fewer sessions might in fact be beneficial, though the authors also admit they could not quite explain this paradoxical result (p. 12) – nevertheless, their findings reinforce the importance of taking into account sessional information.

The second set of variables in Table 2.3 characterises information related to the items (e.g., single-words in lexical retrieval treatments). Words are the raw materials used in language therapies. By extension, details on target words are vital, whether it is for therapists in communication clinics planning interventions (e.g., How many items could be included during the entire course of treatment?) or for aphasia researchers designing intervention studies.

Each variable in the third set (Table 2.4) represents a technique used in therapy. A 'technique' is the particular means utilised by the therapists (it is alternatively referred as

Table 2.2. Information or variables describing the regimen

Variable	Operationalization	Example from Literature
Number of minutes per session Scale of measurement: Ratio	Refers to the average length of time for each session. This provides insights on the temporal space made available for the items within one therapy slot.	Woolf and colleagues (2016): Participants ($n = 5$) were assigned remote therapy to be done at home, using mainstream video conferencing (skype or FaceTime) with the therapist who called from the hospital clinic.
Number of sessions per week Scale of measurement: Ratio	Refers to the average number of sessions per unit of time (laid down as x1 week here for uniformity). This provides insights on how regular sessions are. This is also known as 'dose frequency' in Warren and colleagues (2007).	There were eight sessions with two sessions per week. Each participant completed eight hours of therapy (i.e., each session was an hour). Results suggested that the individuals who completed remote therapy from the hospital clinic, improved in their naming for treated and untreated items.
Number of sessions for the entire therapy Scale of measurement: Ratio	Refers to the total number of sessions provided. This provides insights on how long the entire course of therapy takes.	

Table 2.3. Information or variables describing the items

Variable	Operationalization	Examples from Literature
Total number of items treated across sessions Scale of measurement: Ratio	Refers to the cumulative number of items targeted during the span of therapy. This provides numerical insights on the item coverage.	Snell and colleagues (2010) compared naming performance for the same individuals pre- and post-naming therapy for a 60-item set versus a 20-item set. Although overall analyses in Snell and colleagues' data suggested that regardless of naming deficit severity, participants could tolerate more items, individual analyses revealed that at least two individuals (SS and IH) actually performed better proportionately on smaller sets of items (p. 1087). This highlights the importance of considering the number of items intended for therapy when planning interventions.
Total number of times each item was named across sessions Scale of measurement: Ratio	Refers to the cumulative number of times each item was named during the span of therapy. Also called 'cumulative intervention intensity' in Warren et al.'s vernacular. This provides insights on the amount of naming practice provided for each item.	Harnish and colleagues (2014) tracked the number of naming attempts, and found that at least 2 participants in their sample ($n = 8$) required 1200 naming attempts to achieve significant gains, compared to baseline performance. In other words, the number of times an item was named mattered.
Grammatical class of treated words Scale of measurement: Category – Noun, verb, mixture of nouns and verbs	Refers to whether the target words are nouns, verbs or adjectives etc. Only nouns and verbs (or a mix of these two word classes) happened to be covered by the studies included in this review.	In the initial published trials of Semantic Feature Analysis, the researchers targeted nouns (e.g., Boyle and Coelho, 1995), whereas a similar feature-generation treatment like Verb Network Strengthening

The debate on whether and how words of different grammatical classes are organised or processed differently is a long-standing debate, but beyond the scope of this paper (e.g., see Caramazza and Hillis, 1991, for the theoretical accounts on lexical organisation of nouns versus verbs).

Also, although similar therapies may be applied to target different word classes, it is not clear whether their outcomes would be the same. Combined, this means clinicians should be mindful of the grammatical classes of the words selected for treatment, which could in turn affect other therapy components (e.g., choice of therapy technique(s) etc.).

Treatment (VNeST) targeted verbs as implied by its namesake (Edmonds, 2014).

Early findings supportive of Semantic Feature Analysis were published by the original researchers, with emphatic improvements on treated and untreated items reported (Boyle & Coelho, 1995; see Boyle, 2017, for a review and also how Semantic Feature Analysis had been later modified to accommodate verbs). Recent evidence, however, had been mixed (e.g., van Hees, Angwin, McMahon, and Copland, 2013: Four out of their eight participants did not show item-specific improvement; for five in eight participants, there was no generalisation to untreated items).

Table 2.4. Information or variables describing the techniques

Variable	Operationalization	Examples from Literature
Provision of semantic cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether semantic cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the definition, describing key features of the target item and cloze sentence completion. The cues can be delivered in the spoken and/or written form(s). This variable excludes the semantic cues generated and/or provided during say, Semantic Feature Analysis (SFA). SFA is a multi-step program, distinct from the basic direct application of semantic cues (Boyle & Coelho, 1995).	Abel, Weiller, Huber, and Willmes (2014): One of the therapies investigated involved a hierarchy of semantic cues⁶. The order of semantic cues provided in this particular hierarchy were: (a) the superordinate class of the target word, (b) definition, (c) a closure sentence, and (d) auditory target comprehension (see Abel, Willmes, and Huber, 2007, from which this semantic cue hierarchy was based on). Effects were analysed at the group level ($n = 14$) across therapies (participants also went through therapy using phonological cues) and collectively, there was significant improvement on the treated items when assessed one week after therapy (p. 159).
Provision of phonological cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether phonological cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the first sound or syllable in the spoken form. Compulsory repetition is not required. This variable excludes whole word presentation, which is coded as a separate	Biedermann and Nickels (2008): Therapy relied on presenting cues in progressive order that convey phonological information exclusively. In this case, the cues were: (a) initial phoneme cue (consonant + schwa or vowel) and (b)

⁶ Exactly how a hierarchy of semantic cues should look like is arguably a grey area, with unclear evidence that prioritises one semantic cue over another. In this case, the hierarchy presumably follows the amount of information provided, with the ‘superordinate’ cue being the vaguest or least meaningful.

	variable. It also excludes phonological cues generated and/or provided during Phonological Component Analysis (PCA). PCA is a multi-step program, distinct from the basic direct application of phonological cues (Leonard, Rochon, & Laird, 2008).	tapping the number of syllables in the target word. Participant FME showed significant improvement on treated items (e.g., 'cricket (game)') when assessed one day and one week after therapy. The researchers also distinguished between the types of untreated items. Significant improvement was observed for untreated homophone items (e.g., 'cricket (animal)') when assessed one day and one week after therapy. However, there was no improvement on untreated items that were not homophones (e.g., 'ant')⁷ on both assessment points after therapy.
Provision of orthographic cue(s): Explicit Scale of measurement: Binary – Yes, No	Whether orthographic cue(s) is/are explicitly provided during the actual naming task. Examples include: Presenting the first letter or syllable in the written form. Compulsory copying or reading aloud is not required. This variable excludes whole word presentation, which is coded as a separate variable. It also excludes part-word presentation for copying.	Lorenz and Nickels (2007): In one of their experimental cue conditions, just the initial letter of the target word was presented to facilitate naming. For all three participants, the authors documented instances of significant improvements on treated items when cued with the initial letter, be it assessed immediately or a day after the facilitation session. In one participant, MCB, the effects were even maintained one week later (p. 680).

⁷ The second participant in their study, KCC, showed a slightly different picture, with significantly improved naming on untreated homophones observed one day after treatment, but this was not maintained one week later. Conversely, KCC did not show improved naming on treated items one day after treatment, but showed marginally significant improvement one week later ($p = .083$; see pp. 285-286 in Biedermann and Nickels, 2008, for details). Also note that this is a rare instance in which efforts were made to distinguish between the different types of untreated words.

Provision of phonological cue(s): Implicit Scale of measurement: Binary – Yes, No	Whether phonological cue(s) is/are implicitly provided during the actual naming task. An example includes contextual priming by way of grouping phonologically-related target words together during naming.	Hendricks, Nicolas, and Zipse (2014): In this single-case study, therapy involved having the participant to name phonologically-related words in triplets. Participant AH showed strongest improvement in the ‘front-matched triplets, high phonological neighbourhood density’ treatment condition (post-therapy assessments were done at four time-points: shortly after treatment to 10 weeks afterwards). In this condition, the trio of target words shares the same initial sounds (phonologically-related) and each target belongs to a dense phonological neighbourhood.
Provision of the target word in spoken form Scale of measurement: Binary – Yes, No	Whether the whole target word is provided aurally at any stage of the therapy. Provision of the target word does not necessarily require spoken repetition.	Best, Greenwood, Grassly, and Hickin (2008): The therapy was divided into two parts: (1) A single-word naming therapy, and (2) use of single words in connected speech. The former is relevant here. It included the presentation of the entire word in its spoken form, when the participant ultimately fails to produce it despite progressive cues provided. It is not mandatory for participants to repeat the word. All participants ($n = 8$) showed significant improvement when assessed shortly after the single-word naming therapy (pp. 396-398).

Provision of the target word in written form Scale of measurement: Binary – Yes, No	Whether the whole target word is provided in written form at any stage of the therapy. Provision of the target word does not necessary require reading aloud or copying.	Silkes, Dierkes and Kendall (2013): In this single-case study, therapy involved masked orthographic repetition priming. Written target words were used as the primes, and paired with their corresponding pictures. It is not compulsory to read aloud the words, if seen. Mixed findings were found for the two lists used in treatment. For the ‘animal’ list, no substantial effects were found for both treated and untreated items. For the ‘things to wear’ list, a medium effect size for treated items and a small effect for untreated items were respectively found, when assessed immediately after the treatment (‘A2 vs first 7 baseline points in A1’ in Silkes and colleagues’ Table 3; pp. 390-392).
Spoken repetition of target word: Compulsory Scale of measurement: Binary – Yes, No	Participant is obliged to repeat the given target word verbally, independent of whether the naming response (if required) is correct/incorrect during the therapy. Compulsory repetition can occur at any stage during therapy. The entire target word can be presented in the spoken and/or written form(s).	Mason, Nickels, McDonald, Moses, Makin and Taylor (2011): For therapy, the participants ($n = 3$) were shown a coloured-photograph along with its corresponding sound recording in each trial. The participant was instructed to repeat verbally after the recording. There were no other therapy technique components involved. Two participants (JMM, DRS) did not comply with the therapy instructions. For the participant who complied, SJS, significant improvement was observed for only one set

		of items (Set A) when assessed one week later, and these treatment-specific gains were maintained when assessed against the average baseline scores three weeks later, after the entire course of therapy ended. No generalization to untreated items was found.
Spoken repetition of target word: Contingent on response Scale of measurement: Binary – Yes, No	The participant is only required to repeat the given target word verbally, when the naming response is incorrect. The entire target word can be presented in the spoken and/or written form(s).	Altmann, Hazamy, Carvajal, Benjamin, Rosenbek, and Crosson (2014): In Phases 1 and 2 of the therapy, a picture was presented for naming. For correct responses, the participant will proceed to the next trial. For incorrect responses, the therapist will provide the spoken form of the target word for the participant to verbally repeat. Group analyses ($n = 14$) showed that collectively, compared to baseline picture-naming performance, participants significantly improved on their naming of treated and untreated items when probed immediately post-therapy and three-months later (p. 446)⁸.
Copying of written target word: Compulsory Scale of measurement: Binary – Yes, No	The participant is obliged to copy (orthographically) the given target word, independent of whether the naming response (if required) is correct/incorrect. This	Weill-Chounlamountry, Capelle, Tessier and Pradat-Diehl (2013): In this single-case intervention, part of the therapy involved the compulsory copying of

⁸ There was an additional manipulation, with the 14 participants equally divided into the gesture and no-gesture groups. The only difference is that for the gesture group, a hand movement must be made to initiate the treatment trials (i.e., use the left hand to reach for box and press a red button) or when making naming corrections (left hand must make a non-meaningful circular gesture). This manipulation did not result in any interaction effects (see p. 444, in Altman et al., 2014).

	compulsory act of copying can occur at any stage during therapy. The entire target word is presented in the written form.	the written target word in presence of target picture. Participant HA significantly improved on the naming of trained and untrained words ($n = 90$) when assessed after therapy (time frame not specified, presumably shortly). This improvement was maintained two months later.
Copying of written target word: Contingent on response Scale of measurement: Binary – Yes, No	The participant is only required to copy (orthographically) the given target word, when the naming response is incorrect. The entire target word is presented in the written form.	Wright, Marshall, Wilson, and Page (2008): Therapy included having the participants ($n = 2$) copy down written target words, which (s)he failed to name and write. Participant P2 improved significantly on the naming of all sets of trained items when assessed shortly after treatment and one-month later. Significant improvement was also observed for untreated items when assessed one-month later. Participant P1 only showed significant improvement on the naming of all sets of trained items when assessed shortly after treatment.
Semantic cue(s) generation Scale of measurement: Binary – Yes, No	The participant is required to generate semantic cues. The cues can be generated in spoken and/or written form(s).	Marcotte, Adrover-Roig, Damien, de Préaumont, Gèneux, Hubert, Ansaldo (2012): Therapy involved modified Semantic Feature Analysis. Participants ($n = 9$) were guided to generate semantic features related to the target, e.g., the group it belongs to, its use and location.

		On balance, the nine participants obtained a mean improvement of 80% on the trained items, and 'some degree of generalization ... to untrained materials' (p. 1780).
Phonological cue(s) generation Scale of measurement: Binary – Yes, No	Participant is required to generate phonological cues. The cues can be generated in spoken and/or written form(s).	van Hees, Angwin, McMahon, and Copland (2013): One of the therapies involved phonological component analysis (Leonard et al., 2008). Participants ($n = 8$) were guided to generate the first sound, syllable, last sound, association, and rhyme of the target word. Seven participants improved significantly on the naming of trained items when assessed immediately post-therapy and six maintained the gains two to three weeks later. No generalization was observed to untrained items.
Semantic task(s) Scale of measurement: Binary – Yes, No	The participant performs tasks that focus on the meaning of the target words. Examples include word-to-picture matching task with semantic distractors, semantic categorisation task, or semantic odd-one-out task. In these tasks, the semantic relatedness of the target to other items is typically explored. This variable excludes having the participant generate semantic cues (e.g., Semantic Feature Analysis), as that is coded separately.	Jacquemot, Dupoux, Robotham, and Bachoud-Lévi (2012): Therapy in Phase 1 of this single-case intervention focused on semantics and included semantic categorisation (labelled as 'picture categorisation' in their paper), function-matching (selecting the picture that best represents the function of the target item), semantic-verification (labelled as 'knowledge assessment') tasks.

		Participant DPI improved on his naming of the trained items after therapy in Phase 1, but this improvement was not significant.
Orthographic task(s) Scale of measurement: Binary – Yes, No	The participant performs tasks that focus on the spellings of the target words. Examples include anagrams, first letter or syllable identification. In these tasks, letter identification and letter sequencing of the target word are typically explored.	Weill-Chounlamountry, Capelle, Tessier and Pradat-Diehl (2013): Part of the therapy involved the rearranging jumbled letters in this single-case study. Participant HA significantly improved on the naming of trained and untrained words. This improvement was maintained two months later.

Table 2.5. Information or variables describing the application of technique(s)

Variable	Operationalization	Examples from Literature
Application of cues: Progression – Increasing vs. Decreasing Scale of measurement: Category – (1) Increasing, (2) Decreasing, (3) Not applicable	Whether the part-word cues are applied in an increasing (from small to bigger units; e.g., single phoneme to syllable) or decreasing (from big to smaller units; e.g., syllable to single phoneme) progression.	Conroy, Sage, and Lambon Ralph (2009d) compared the two approaches of cueing (increasing versus decreasing). At the individual level, all participants ($n = 7$) benefitted from both approaches, as each approach yielded significant improvements on naming for their respective set of treated items. The significant improvements were observed when assessed immediately at one-week post-therapy and at follow-up, five weeks after therapy ended.
Application of cues: Contingent on response vs. compulsory regardless of response (shortened as ‘Application of cues: Contingent on response’ in Figures 2 to 3) Scale of measurement: Category – (1) Contingent, (2) Compulsory application, (3) Mixture of both, (4) Not applicable	Whether the cues are applied according to the response or independent of the response.	In Biedermann and Nickels (2008; also reported in Table 4), whether the cues were applied depended on the accuracy of the response. Cues were progressively applied from initial phoneme to whole word repetition, if responses were incorrect or not made within 10 seconds. Harnish and colleagues (2014) illustrated the obligatory application of cues irrespective of naming accuracy. There were eight steps (including different cues at different times) in their therapy, which ‘progressed through each of the eight presentations regardless of the participants’ responses ... This provided opportunity for multiple repetitions of the correct responses.’ (p. S288). The authors argued that all participants ($n = 8$) showed

		treatment-specific gains in naming post-therapy, and out of the seven assessed 8 weeks after follow-up, six maintained their gains (see p. S291 in their original paper). There was no reliable evidence of generalization to untreated items. Conroy and Scowcroft (2012) is an example of a mixture of cues. The cueing process began with the compulsory spoken repetition of the whole word and then it became contingent on the participant's response – i.e., if the participant failed to name when picture was presented for a second time, the whole word was presented again, else the experimenter proceeded to the next round using consonant-vowel (CV) syllable as cues (see p. 307 of their paper for a description). All participants ($n = 4$) in Conroy and Scowcroft (2012) showed significant improvements on their naming of treated items post-therapy, and these were also significantly higher than scores posted for the untreated control items.
Feedback on accuracy of the naming response Scale of measurement: Category – (1) No feedback, (2) explicit feedback provided, (3) implicit feedback provided	The type of feedback provided to the participant about naming accuracy. This is for each of the naming response made during therapy. There can be (1) no feedback, (2) explicit feedback, or (3) implicit feedback.	Off, Griffin, Spencer, and Rogers (2016) is an example of a therapy with nil feedback. 'Participants did not receive feedback at any time during training. That is, participants did not have information about whether they accurately produced the correct name for the target.' (p. 993)

	Explicit feedback refers to instances in which at all stages of therapy where naming responses are needed, there is direct acknowledgement of accurate responses and pointing out of inaccurate responses. Implicit feedback refers to instances in which at any stage in the therapy where naming responses are needed, direct acknowledgement of the accuracy of response is not present, but can be inferred. Typically, this takes the form of progressing to next naming trial only if the response is correct. Cues, however, are provided if the response is incorrect.	Macoir, Routhier, Simard and Picard (2012) is an example of a therapy with explicit feedback. 'Feedback was given for each response ... When an error was produced, the patient was given the correct answer, along with an explanation about the nature of the error.' (p. 124) Best and colleagues (2008; also reported in Table 4) is an instance of a therapy with implicit feedback. Although feedback was not directly provided to the participants on the accuracy of their responses, they could infer based on whether the experimenter/therapist proceeded to the next trial or supplied them with a cue.
Termination criteria Scale of measurement: Category – (1) Performance criterion; (2) Fixed number of sessions completed; (3) Mix of performance criterion and certain number of sessions completed	Therapy can be broadly terminated in three ways: (1) The participant manages to achieve a certain level of performance (i.e., criteria present); (2) The participant has completed a fixed number of sessions (i.e., criteria absent); or (3) The participant meets a certain level of performance or manages to complete a fixed number of sessions (i.e., mix of performance criterion and a fixed number of sessions)	Rider, Wright, Marshall, and Page (2008) is an example of a therapy with performance criteria. The participant moved on to the next stage of the therapy (i.e., names next list of items in this case), after achieving 80% naming accuracy across two consecutive trials, or having completed 10 trials (p. 164). Conroy, Sage, and Lambon Ralph (2009a) is an example of a therapy without performance criteria. The participants were simply required to complete 10 sessions of therapy (p. 474).

		In Marcotte, Adrover-Roig, Damien, de Preaumont, Genereux, Hubert, and Ansaldo (2012), therapy was terminated when participants achieved 80% performance on treated items or completed a maximum of six weeks of therapy (p. 1777).
Mode in which the actual therapy was delivered Scale of measurement: Category - (1) Machine, (2) Individual, (3) Combination of machine and individual	Whether the actual therapy is delivered by: (1) the machine (i.e., computer-based/assisted or self-administered therapy such that the computer is providing the cues, for instance; the use of computer can be unsupervised, supervised or assisted), (2) a person (the therapist / experimenter/ trained individual is providing the cues) or (3) a combination of machine and person (e.g., orthographic cues presented on the computer, but phonological cues are supplied by the therapist; or some parts of therapy are completely computerised but other parts are solely provided by the therapist)	Adrian, Gonzalez, Buiza, and Sage (2011) pilot-tested the Spanish Computer-assisted Anomia Rehabilitation Program (CARP). It is an example of a computer providing the actual therapy cues. All their participants ($n = 15$) improved significantly on the naming of treated items and 11 of them showed generalization to untreated item post-treatment (seven to ten days after the treatment stage; also see Routhier, Bier, and Macoir, 2016, for a self-administered word-finding therapy on a tablet). Harnish and colleagues (2014; see above) is an example of a therapy that uses a mixture of both computer and therapist to provide therapy. The orthographic cues were presented on the computer screen, but the phonological and semantic cues were spoken aloud by the therapist (see p. S288 in the original paper).
Homework Scale of measurement: Binary – Yes, No	Whether homework is assigned as part of the therapy package	Conroy and Scowcroft (2012) is an example in which homework was assigned. Participants ($n = 4$; likely with assistance of caregivers) were asked to look at the pictures and read aloud the corresponding written words underneath three times.

Unfortunately, despite explicit instructions requesting participants to document their frequencies of homework completion, no participants documented so with sufficient consistency to permit analysis on the relationship between homework and naming progress (see p. 315 of the original paper).

‘approach’ by Brady and colleagues, 2016). Variables in the third set are considered the core of the therapy. To illustrate, these are some potential ‘techniques’ relevant for treating naming difficulties secondary to impaired phonological output processes: (a) Getting clients to generate phonological features (e.g., in Phonological Component Analysis); (b) completing phonological tasks (e.g., rhyme judgment); or (c) confrontational naming assisted by part-word phonological and orthographic cues, to list a few. In clinical practice, these techniques are not mutually exclusive and can be jointly applied with proper rationales.

The last set of variables listed in Table 2.5 pertains to the applications of technique(s) and adjuncts. They either elaborate on how techniques are executed (e.g., increasing or decreasing progression of cues; Conroy, Sage, & Lambon Ralph, 2009d) or lay down general principles about the process (e.g., the termination criteria that specifies when to cease; also see Howard, Best, & Nickels, 2015, pp. 533-534). Adjuncts that could complement therapy, like homework, are also included here.

2.1.3 Random forest and its utility in word-finding therapy research

The multiple components making up the therapies present analysts with a challenge: How can the individual effects of these parts be meaningfully disambiguated? Conventional ways to conduct meta-analysis fall short. They target to calculate a pooled effect size, in order to quantitatively summarise findings obtained from a body of empirical research. Consequently, to quote, ‘the tidiest meta-analysis result in this regard is a distribution of effect sizes that are relatively homogenous under the fixed effects model ... [However] When the distribution of effect sizes is heterogenous ... the average over contrary results is not likely to converge on the truth, just muddle it.’ (Lipsey and Wilson, 2001, p. 162; also see Hall and Rosenthal, 1995, for criticisms on traditional approaches to meta-analysis when studies are heterogenous and complex). In the context of aphasia intervention (to which word-finding intervention belongs), discussion of heterogeneity is relevant, due to the diversity in implementations as well as patient characteristics (e.g., in Wisenburn and Mahoney (2009), the *SD* for effect size *d* of each therapy approach was ‘extremely high’ (p. 1347) and ranged from ‘1.23’ to ‘4.12’ (p. 1344) as acknowledged by the authors; the same brain lesion may also result in different cognitive/language impairments)⁹. Clearly, what is needed is an alternative means of analysing the wealth of word-finding intervention data without falling victim to the illusion of homogeneity.

⁹ Unfortunately, in the Cochrane review by Brady and colleagues (2016), the heterogeneity index, *I*² statistic, was not reported.

Random forest is an established machine learning statistical technique that leverages the power of multiple decision trees to identify robust predictors. A full account of random forest is not within the scope of this thesis. Interested readers are referred to the original work by Breiman (2001) and his accompanying website¹⁰ for the finer details. A rudimentary exposition to facilitate subsequent understanding is however, provided: From an original dataset, bootstrap samples (or training sets, using Breiman's terminology) are randomly drawn. For each bootstrap sample, a unique decision tree is constructed using the principles of 'Classification and Regression Trees (CART)' (Morgan & Sonquist, 1963; Breiman, Friedman, Olshen & Stone, 1984), with the nodes split by the best predictor among a subset of variables randomly chosen for that node. Multiple decision trees, each grown based on these rules, assemble to form a random forest.

Random forest is deemed suitable for our purposes because: Firstly, it has the versatility to handle data from different designs, including single case studies (see de Aguiar and colleagues, 2016, for an example). On the other hand, classical meta-analytic approaches do not lend themselves easily to single case research: Their typical formulae set out to summarise effect sizes extracted from group studies, not single case studies (specifically, they required group mean differences and standard deviations at the group level to be entered as input, see Borenstein, Hedges, Higgins, and Rothstein, 2009).

Secondly, a main strength of random forest is its ability to manage 'wide data', i.e., there is a large number of variables but relatively few observations (i.e., 'small n , large p ' or 'high dimension, low sample size' problems; see Bureau and colleagues, 2005, for a demonstration). As our meta-analysis (elaborated later) relies on a modest sample of observations (maximum n for an outcome = 184) while seeking to examine the roles of 31 variables, random forest appears particularly suitable.

In addition, the variables examined here are a mixture of categorical and continuous scales of measurement. Unlike regression (meta-regression being the case in point), this does not pose a problem for random forest, which is sufficiently robust to jointly analyse these variables with assorted scales of measurement together.

Finally, random forest produces an output quantifying the relative importance of the variable (i.e., 'variable importance'). This allows us to gainfully isolate and distinguish key variables (e.g., a particular therapy component) responsible for specific treatment outcomes (e.g., perhaps therapies using phonological cues lead to successful lexical retrieval of treated items within three weeks post-therapy, but not after three weeks). Notably, variable importance of a predictor is innovatively computed by measuring the loss in predictive

¹⁰ https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

accuracy if its original values are substituted through permutation (Brieman, 2001; also see Hothorn, Hornik, Strobl, & Zeileis, 2019, p. 33). Conceptually, this means that during evaluation the predictor is always weighed against the entire host of variables included in the forest (i.e., the rest of the variables will have their individual predictive utilities accounted for, and they are not simply treated as random noise). This is an asset in situations when appraisal of multiple variables is required. The superiority of random forests over other statistical approaches to locate predictors is well-documented across disciplines: For example, Tagliamonte and Baayen (2012) showed how random forest could ‘overcome the limitations of mixed-effects models’ (p. 165) using sociolinguistics data.

2.1.4 Purpose of this systematic review and meta-analysis

To recap, there is a lack of research that formally breaks single-word spoken naming therapy down into its multiple components for scrutiny. A novel analysis that integrates existing research findings to evaluate these various therapy components is urgently needed. The overall objective of the first part of this chapter is therefore to review and synthesise recent evidence obtained for language-based approaches (i.e., semantic, phonological and orthographic techniques). This collection of empirical evidence will be based on interventions designed to facilitate single-word retrieval in spoken word production for adults with impaired naming after a stroke. Papers published between January 2008 and May 2018 are included. This will allow us to provide an update based on the information generated roughly within a decade.

In particular, the interest here is in determining major therapy component(s) that underpin treatment efficacy, and whether the component(s) differ for treated and untreated items when assessed shortly or at longer-term periods after the therapy had terminated. Short-term is operationalised as being within three weeks post-therapy, and maintenance phase as being four weeks and more post-therapy. While there are no ‘official’ benchmarks for these time frames, these operationalisations are consistent with how other studies have defined these periods (e.g., Croft, Marshall, Pring & Hardwick, 2011). Finally, random forest, given its flexibility, strengths and capacity to penetrate assorted data, is harnessed to run the meta-analysis and uncover the ‘active’ therapy components.

2.2 Method

2.2.1 Information sources and search

A systematic search was conducted using a combination of electronic databases, trial registers, and grey literature. In total, 17 sources (see Appendix 2.1) were involved.

The keywords used in the search included stroke/cerebrovascular accident/cerebral ischemic/cerebral haemorrhage AND aphasia/dysphasia/anomia/naming deficit/naming impairment/naming difficulty/naming problem/word finding deficit/word finding impairment/word finding difficulty/word finding problem/word retrieval deficit/word retrieval impairment/word retrieval difficulty/word retrieval problem AND therapy/intervention/cue/facilitation AND adult. Appropriate modifications were made according to the search engine's requirements. Finally, the search included papers published from January 2008 to May 2018. Based on the procedure described, a total of 5,650 entries were generated (Figure 2.1).

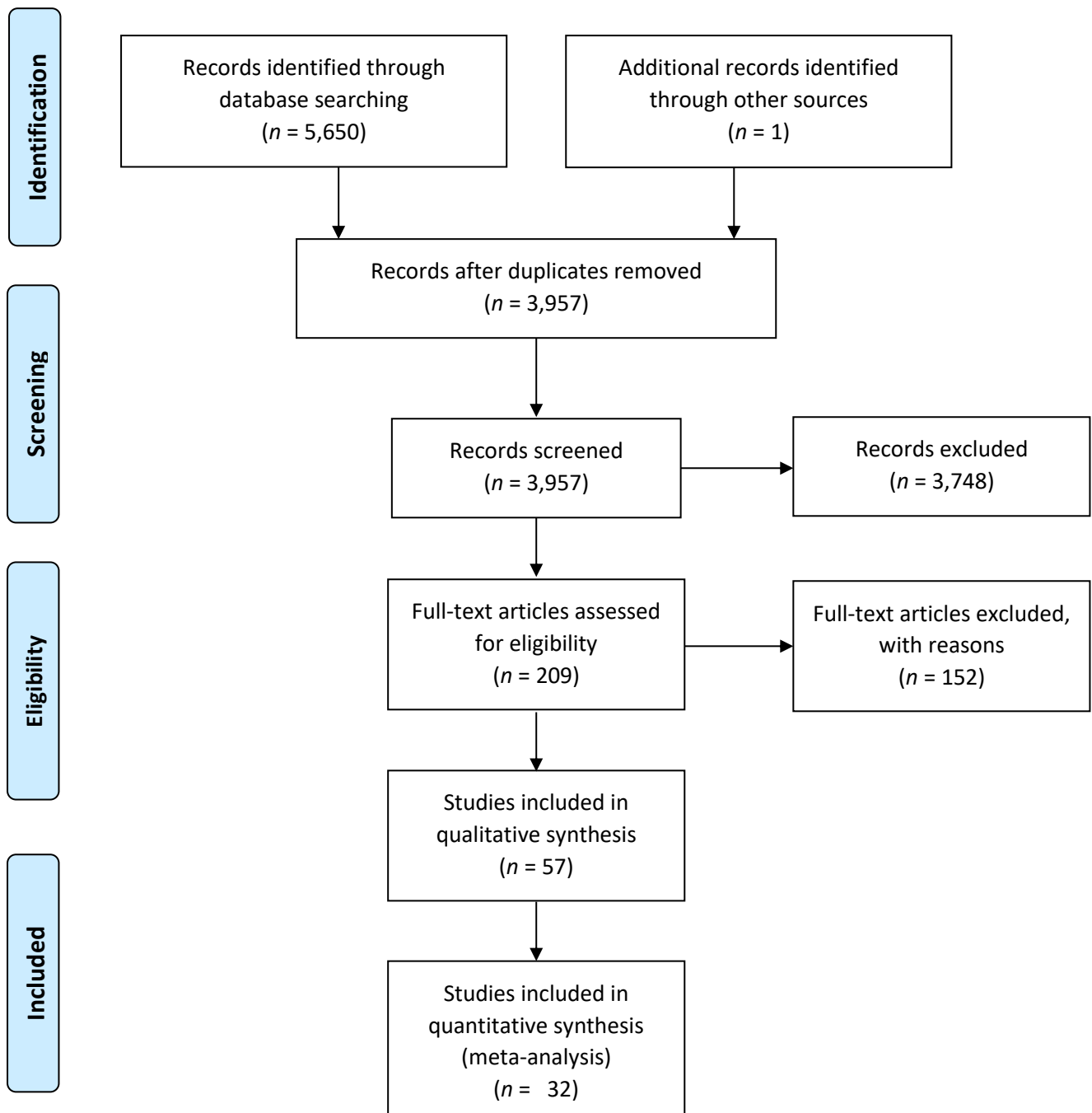
2.2.2 Study selection and eligibility criteria

After removing duplicates, the author and an external reviewer independently inspected the 3,957 entries against the following pre-specified criteria. The studies must be single-word naming therapy research published in peer-reviewed journals or non-published theses as long as they were written in English and contained original data (i.e., research protocols without any published data are excluded, unless their data are included in the trial registers). Naming therapies conducted in non-alphabetic languages were excluded because differences in language properties could potentially necessitate different treatment requirements during naming therapies (e.g., the prevalence of homophony in say, Mandarin, may possibly limit the utility of phonological cues). Only studies using therapy techniques targeting single-word naming, and involving language-based facilitation by way of semantic, phonological, and/or orthographic approaches were included. Studies investigating the use of constraint-induced therapy, gestural cues, tactile cues or acoustic environmental sounds as cues were excluded.

In terms of experimental designs, both group designs (these may be randomised control trials (RCTs) or not), and single case experimental designs (i.e., case series and single case studies) were included. Single case experimental designs were specifically included because in aphasia research, they constitute a substantial portion of the published evidence¹¹. In terms of outcome variables, only studies whose single-word accuracy outcomes were obtained using the spoken picture naming paradigm, were included.

¹¹ A rudimentary count revealed that for the journal 'Aphasiology', single case or case series intervention studies made up 20% of the original articles published in 2018.

Figure 2.1. PRISMA flow chart for systematic review (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).



Studies with the following outcomes were thus excluded: (a) Outcomes that do not involve accuracy and/or naming (e.g., quality of life ratings); (b) accuracy outcomes that are non-verbal (e.g., gestural or written output); and (c) accuracy outcomes that are measured at the phrase or sentence level.

In terms of participants, studies that recruited adults with aphasia after stroke, specifically those with spoken naming difficulties, were included. Not included were data collected from adults whose language impairment did not concern naming (e.g., impaired comprehension). Also excluded were studies recruiting from adults whose naming difficulties stemmed from other acquired deficits (e.g., traumatic brain injury) or neurodegenerative disorders (e.g., primary progressive aphasia).

2.2.3 Data extraction for meta-analysis

A total of 57 studies were identified and jointly agreed by two reviewers independently as fulfilling the criteria stated. The screening process was first completed using abstracts and later, the full articles. Discrepancies between the two reviewers were settled through online face-to-face discussions. Interrater reliability between the reviewers was strong ($\kappa = 0.80$). Out of the 57 studies, 32 papers (Appendix 2.2) reported therapy-specific information at the individual participant level, and are therefore eligible for data extraction. From these papers, a dataset containing 222 participant cases was compiled. An external clinician was solicited to further independently verify around 10% of the extracted data (randomly selected); no discrepancies were found.

The dataset comprised 31 independent variables spread across five categories. Each variable was extracted at the participant-level from individual papers.

Outcome measures. Separate outcomes were extracted for treated and untreated items respectively at the short-term and maintenance phases, generating four outcome measures. To recap, short-term was operationalised as being within three weeks post-therapy, and maintenance phase as being four weeks and more post-therapy. The outcome measures were binary coded, with '1' indicating 'improved naming performance'. To qualify as 'improved', the result must either show a significantly better naming performance as compared to baseline using statistical analyses, or report an effect size of at least a value of '4'. The latter observes the benchmarks set in Beeson and Robey (2006, p. 167), which are tailored for lexical-retrieval treatments in single-case studies. Case reports whose results were inferred solely using visual analyses, e.g., split-middle lines (Morley, 2018, pp. 99-101), were excluded.

There are unfortunately no definitive guidelines on optimal numbers necessary for a random forest classifier to operate. It has been tentatively recommended for datasets to contain at least 100 observations when using general CART programs (Steinberg, n.d.). To be conservative, we pitched a minimal threshold of 120 observations for 31 predictors. For this reason, the analysis of untreated items at maintenance was dropped, as its number of observations fell below the threshold ($n = 105$). This leaves three outcome measures for our meta-analysis, i.e., naming performance for treated items at the short-term and maintenance phases, as well as naming performance for untreated items at the short-term phase.

Independent variables. The five categories of independent variables are: (1) Participant-related variables, (2) therapy-related variables about the regimen, (3) therapy-related variables about the items (i.e., target words); (4) therapy-related variables about the technique(s); and (5) therapy-related variables about the application of technique(s) and their adjuncts. Participant-related variables describing the demographic and clinical information of each individual were: the participant's (a) age (in years), (b) gender (male/female), (c) years of formal education received, (d) the number of months post stroke-onset and (e) baseline naming scores according to the 'Boston Naming Test' (Kaplan, Goodglass, & Weintraub, 2001). These participant-related variables are included as they represent standard control factors. 'Boston Naming Test' was chosen to index baseline naming performance as it emerged to be the popular choice in the papers covered in this review (140 out of 222 possible cases (approximately 63.60%) used the 'Boston Naming Test' to assess initial picture naming performance; whereas 76 cases (34.23%) used another standardised instrument like the 'Object and Action Naming Battery' (OANB; Druks & Masterson, 2000)).

The rest of the predictor variables are covered in depth earlier (see 'Therapy components or "active ingredients"; i.e., Tables 1.2 to 1.5). It is these therapy-related variables, specifying information about what constitutes the treatments that this thesis is primarily interested in.

Of note, variables specifying participants' language deficits identified by way of baseline assessments, were not included within our set of predictors. This is due to practical concerns over reliability and heterogeneity: Across the studies reviewed, there is no uniformity in the types of assessments selected by researchers. The assessments could be various standardised tests either used in their complete formats or abridged for expediency, or unstandardised measures with unknown psychometric properties, often devised specifically for the projects (e.g., a bespoke 'definition-to-written word matching task' in Macoir, Routhier, Simard & Picard, 2012). When compiled, there is substantial heterogeneity, which does not make for a clean variable. Variables pertaining to language

deficits deduced by baseline assessments were thus excluded as predictors in the subsequent analyses.

2.2.4 Meta-analysis

The participant and therapy-related variables were modelled for each of the three outcome measures separately using random forest. Random forest is a supervised machine learning method. Through constructions of multiple decision trees, information is classified. The aggregated data yielded from a forest of trees is then able to provide stable and accurate predictions (Breiman, 2001). All data was processed in R (R Core Team, 2019) based on the steps described below. The implementation and programming were informed by Breiman (1999), Strobl, Hothorn, and Zeileis (2009), Tagliamonte and Baayen (2012), and de Aguiar and colleagues (2016).

Procedure. First, missing values were imputed following de Aguiar and colleagues (2016) in three steps: (a) Variables missing 56% of their raw information (e.g., ‘type of stroke (ischemia vs haemorrhage)’ was missing 71.62 % of its data) were excluded (rationale explained below). (b) The `rflmpute` function in ‘randomForest’ was then used to impute the remaining missing values. To obtain the estimates, 100 iterations of 2000 trees were run. The missing values were imputed and adjusted using the proximity matrix generated by random forest (see Liaw & Wiener, 2018). (c) Finally, these steps were repeated 20 times to generate 20 different databases. These numbers follow figures cited in literature: For instance, Shah, Barlett, Carpenter, Nicholas, and Hemingway (2014) conducted imputation simulations on variables with a maximum of 56.8% missingness, and recommended running more than 10 repetitions for data with comparable missingness (p. 772). Similarly, de Aguiar and colleagues were confident and ascertained their results to be reliable after running 20 repetitions (but see Wei et al., 2010, who was able to address model stability by conducting only four random forest simulations per dataset). To be conservative and in keeping with de Aguiar and colleagues, 20 databases were similarly simulated here, to facilitate subsequent ‘variable importance’ computations.

Next, the ‘`cforest` function’ was used to construct a random forest for each database (Hothorn, Hornik, Strobl, & Zeileis, 2019). ‘`cforest`’ is an enhanced version of random forest, and corrects for multicollinearity between variables to produce unbiased variable selection for each decision tree (see Strobl, Boulesteix, Zeileis & Hothorn, 2007, who ran simulations to show that their usage of conditional permutation importance within ‘`cforest`’ was able to protect against bias arising from multicollinearity). To optimise performance, we also followed the creators’ recommendations and set the number of predictors randomly sampled

at each node to be '5' (p. 17, Strobl, Hothorn, & Zeileis, 2009) and the number of trees grown to be '5000'. In addition, to assess classification accuracy, the concordance statistic was computed using 'treeresponse' command as per specified in Tagliamonte and Baayen (2012), and the 'Out Of Bag' (OOB) error for each random forest extracted using 'caret' (Kuhn, 2008, 2018). The concordance index provides a gross indication of the predictive ability of the classification generated by the random forest model. Based on extant literature, a concordance index of '0.5' indicates 'random performance', but a value exceeding '0.83' a 'strong classification model' (Hermansen, 2008). OOB error, on the other hand, is a statistic innate to random forest. It measures classification accuracy through a built-in model-testing against a unique holdout sample (subsample) not used in the training of that decision tree. A comparable OOB estimate is a maximum threshold of '0.18', based on de Aguiar and colleagues (2016).

In the final step, the random forest most representative of the overall findings was identified. This was accomplished in two steps for each outcome measure: (a) For every database, the variables were ranked according to their respective values of variable importance. An overall median rank was computed across the twenty databases for each variable. (b) The rankings of each database were then compared against the overall median ranks using Kendall's *Tau-b*. The latter is appropriate as it examines the compatibility of two sets of rankings by taking into account the number of paired observations that are concordant and discordant (Kendall, 1938; pp. 188-189 in Agresti, 2010)¹². The database with the highest Kendall's rank coefficient was thus chosen to be the representative random forest for final interpretation.

2.3 Results

Based on the 222 eligible cases, results for the three outcome measures were computed, and they are reported individually below. The overall descriptive statistics for missing data are reported in percentages as follow: There were 45.33% values missing for 'number of minutes per session', 39.11% for 'total number of times each item was named across sessions', 35.56% for 'number of sessions per week', 33.78% for 'Boston Naming Test score at baseline', 28.89% for 'grammatical class of treated words', 25.33% for 'years of education received', 22.22% for 'provision of the target word in written form', 14.67% for

¹² Although not stated in the analyses, for each database, we created scatterplots between the median rankings across 20 databases and the individual database's rankings to ensure that in each of them, the relationship between the two types of rankings was monotonic. All scatterplots showed a linear distribution in the increasing direction. Monotonicity is an assumption in Kendall's *Tau-b*.

'mode in which therapy was delivered', 12% for 'feedback per naming response', and 10.22% for the 'total number of sessions'. The remaining predictors had less than 3.5% values missing; there were no missing values for 14 variables.

2.3.1 Successful treatment outcome of treated items post-therapy in the short-term phase

Within the 222 possible cases, 184 cases reported data for the individual's spoken naming performance on treated items within three weeks after therapy; out of which, 167 instances documented improvement and 17 showed no difference in their naming performance pre- and post-therapy (see Table 2.6 for descriptive statistics; note the descriptive statistics in this section are based on the respective representative random forest *after* the missing values had been imputed). The representative random forest produced a concordance statistic [$C = 0.968$] and an OOB error rate of 0.092. Both diagnostic figures fall well within the aforementioned bandwidth of good performance. These suggest that the imputed random forest classification is a strong predictive model. Kendell's *Tau-b* coefficient showed that the rankings correlated highly and significantly with the median rankings obtained across the 20 databases [$r_T = .984, p < .0001$]. Figure 2.2 summarises our main findings based on the representative database for this measure, with the variables arranged in order of their importance. In descending order, the top three variables associated with improvement for naming treated items in the short term after therapy were: (1) Provision of the target word in the written form; (2) Explicit provision of orthographic cue(s), i.e., part-word cues like initial letter and syllables; and (3) Application of cues: Whether the application is contingent on response or compulsory regardless of response.

2.3.2 Successful treatment outcome of treated items post-therapy during the maintenance phase

In total, 127 identifiable cases provided individual-level spoken naming data on treated items in a period beyond three weeks after therapy; and 116 of these reported improvements (when compared to baseline), whilst 11 detected no difference (see Table 2.7). The representative random forest also fulfilled the diagnostic requirements of good performance [concordance index $C = 0.962$; OOB = 0.087]. High and significant correlation was also found between the model's rankings and the overall median rankings [$r_T = .941, p < .0001$]. The main findings related to variable importance are reported in Figure 2.3, with (1) Feedback on accuracy of naming response, (2) Provision of the target word in the written form, and (3) Explicit provision of orthographic cue(s), found to be the three best predictors.

Table 2.6. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on treated items, when assessed at the short-term phase after therapy.

Name of variable	Mean (SD) or frequency count of cases, where applicable	
	Improved outcome observed (<i>n</i> = 167)	Nil improvement in outcome observed (<i>n</i> = 17)
1. Participant-related variables		
Age (Years)	61.35 (<i>SD</i> = 11.62)	56.12 (<i>SD</i> = 11.71)
Gender	Male: 88; Female: 79	Male: 10; Female: 7
Months post-onset of stroke	41.29 (<i>SD</i> = 45.26)	76.53 (<i>SD</i> = 102.24)
Years of education received	11.87 (<i>SD</i> = 3.48)	11.93 (<i>SD</i> = 1.98)
Boston Naming Test score at baseline (Percent)	0.41 (<i>SD</i> = 0.21)	0.49 (<i>SD</i> = 0.25)
2. Variables describing the regimen		
Number of minutes per session	56.63 (<i>SD</i> = 7.98)	63.84 (<i>SD</i> = 12.45)
Number of sessions per week	2.17 (<i>SD</i> = 0.61)	2.97 (<i>SD</i> = 0.59)
Number of sessions for the entire therapy	13.94 (<i>SD</i> = 7.80)	12.71 (<i>SD</i> = 3.30)
3. Variables describing the items (target words)		
Total number of items treated across sessions	46.12 (<i>SD</i> = 42.93)	38.24 (<i>SD</i> = 13.85)
Total number of times each item was named across sessions	70.66 (<i>SD</i> = 27.53)	31.87 (<i>SD</i> = 15.79)
Grammatical class of the treated words	Nouns: 42; Verbs: 64; Mixture of nouns and verbs: 61	Nouns: 5; Verbs: 10; Mixture of nouns and verbs: 2
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 97; No: 70	Yes: 5; No: 12
Provision of phonological cue(s): Explicit	Yes: 145; No: 22	Yes: 6; No: 11
Provision of orthographic cue(s): Explicit	Yes: 149; No: 18	Yes: 2; No: 15

Provision of phonological cue(s): Implicit	Yes: 1; No: 166	No: 17
Provision of the target word in spoken form	Yes: 167	Yes: 13; No: 4
Provision of the target word in written form	Yes: 151; No: 16	Yes: 2; No: 15
Spoken repetition of target word: Compulsory	Yes: 55; No: 112	Yes: 1; No: 16
Spoken repetition of target word: Contingent on response	Yes: 147; No: 20	Yes: 13; No: 4
Copying of written target word: Compulsory	Yes: 1; No: 166	No: 17
Copying of written target word: Contingent on response	No: 167	No: 17
Semantic cue(s) generation	Yes: 3; No: 164	Yes: 5; No: 12
Phonological cue(s) generation	Yes: 6; No: 161	Yes: 2; No: 15
Semantic task(s)	Yes: 32; No: 135	No: 17
Orthographic task(s)	Yes: 1; No: 166	No: 17
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 102; Decreasing: 36; Not applicable: 29	Increasing: 5; Not applicable: 12
Application of cues: Contingent on response	Contingent on response: 85; Compulsory: 19; Mixture of compulsory and response-contingent applications: 52; Not applicable: 11	Contingent on response: 5; Compulsory: 1; Not applicable: 11
Feedback on accuracy of the naming response	Explicit feedback: 6; Implicit feedback: 153; No feedback: 8	Explicit feedback: 4; Implicit feedback: 6; No feedback: 7
Termination criteria	Performance criterion: 1; Fixed number of sessions: 166	Fixed number of sessions: 17
Mode in which actual therapy was delivered	Computer: 62; Person: 105	Computer: 2; Person: 15
Homework	Yes: 8; No: 159	No: 17

Table 2.7. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on treated items, when assessed during the maintenance phase after therapy.

Name of variable	Mean (SD) or frequency count of cases, where applicable	
	Improved outcome observed (<i>n</i> = 116)	Nil improvement in outcome observed (<i>n</i> = 11)
1. Participant-related variables		
Age (Years)	62.63 (<i>SD</i> = 11.44)	59.55 (<i>SD</i> = 14.36)
Gender	Male: 69; Female: 47	Male: 10; Female: 1
Months post-onset of stroke	48.90 (<i>SD</i> = 60.43)	133.45 (<i>SD</i> = 177.24)
Years of education received	13.43 (<i>SD</i> = 1.58)	13.49 (<i>SD</i> = 1.86)
Boston Naming Test score at baseline (Percent)	0.40 (<i>SD</i> = 0.22)	0.33 (<i>SD</i> = 0.11)
2. Variables describing the regimen		
Number of minutes per session	53.65 (<i>SD</i> = 8.40)	61.82 (<i>SD</i> = 14.71)
Number of sessions per week	2.18 (<i>SD</i> = 0.66)	3.15 (<i>SD</i> = 0.62)
Number of sessions for the entire therapy	10.30 (<i>SD</i> = 2.29)	12.09 (<i>SD</i> = 4.29)
3. Variables describing the items (target words)		
Total number of items treated across sessions	30.85 (<i>SD</i> = 15.75)	22.91 (<i>SD</i> = 14.52)
Total number of times each item was named across sessions	69.03 (<i>SD</i> = 29.12)	51.50 (<i>SD</i> = 21.32)
Grammatical class of the treated words	Nouns: 26; Verbs: 40; Mixture of nouns and verbs: 50	Nouns: 6; Verbs: 4; Mixture of nouns and verbs: 1
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 49; No: 67	Yes: 1; No: 10
Provision of phonological cue(s): Explicit	Yes: 97; No: 19	Yes: 2; No: 9
Provision of orthographic cue(s): Explicit	Yes: 107; No: 9	Yes: 3; No: 8

Provision of phonological cue(s): Implicit	Yes: 1; No: 115	No: 11
Provision of the target word in spoken form	Yes: 115; No: 1	Yes: 7; No: 4
Provision of the target word in written form	Yes: 111; No: 5	Yes: 3; No: 8
Spoken repetition of target word: Compulsory	Yes: 50; No: 66	Yes: 2; No: 9
Spoken repetition of target word: Contingent on response	Yes: 91; No: 25	Yes: 5; No: 6
Copying of written target word: Compulsory	Yes: 1; No: 115	No: 11
Copying of written target word: Contingent on response	Yes: 1; No: 115	No: 11
Semantic cue(s) generation	Yes: 3; No: 113	Yes: 1; No: 10
Phonological cue(s) generation	No: 116	No: 11
Semantic task(s)	No: 116	No: 11
Orthographic task(s)	Yes: 2; No: 114	No: 11
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 63; Decreasing: 29; Not applicable: 24	Increasing: 3; Not applicable: 8
Application of cues: Contingent on response	Contingent on response: 83; Compulsory: 18; Mixture of compulsory and response-contingent applications: 8; Not applicable: 7	Contingent on response: 3; Compulsory: 1; Not applicable: 7
Feedback on accuracy of the naming response	Explicit feedback: 1; Implicit feedback: 109; No feedback: 6	Explicit feedback: 3; Implicit feedback: 2; No feedback: 6
Termination criteria	Performance criterion: 1; Fixed number of sessions: 109; Mixture of the two: 6	Fixed number of sessions: 9 Mixture of the two: 2
Mode in which actual therapy was delivered	Computer: 6; Person: 110	Computer: 3; Person: 8
Homework	Yes: 8; No: 108	No: 11

Table 2.8. Descriptive statistics based on imputed values generated by the representative random forest for spoken naming performance on untreated items, when assessed at the short-term phase after therapy.

Name of variable	Mean (SD) or frequency count of cases, where applicable	
	Improved outcome observed (<i>n</i> = 27)	Nil improvement in outcome observed (<i>n</i> = 125)
1. Participant-related variables		
Age (Years)	59.67 (<i>SD</i> = 13.88)	62.25 (<i>SD</i> = 11.12)
Gender	Male: 9; Female: 18	Male: 74; Female: 51
Months post-onset of stroke	31.63 (<i>SD</i> = 25.05)	48.20 (<i>SD</i> = 61.14)
Years of education received	9.14 (<i>SD</i> = 5.40)	12.55 (<i>SD</i> = 2.72)
Boston Naming Test score at baseline (Percent)	0.40 (<i>SD</i> = 0.11)	0.43 (<i>SD</i> = 0.23)
2. Variables describing the regimen		
Number of minutes per session	58.29 (<i>SD</i> = 4.67)	59.81 (<i>SD</i> = 8.35)
Number of sessions per week	2.30 (<i>SD</i> = 0.22)	2.35 (<i>SD</i> = 0.55)
Number of sessions for the entire therapy	26.30 (<i>SD</i> = 7.92)	11.80 (<i>SD</i> = 5.16)
3. Variables describing the items (target words)		
Total number of items treated across sessions	85.19 (<i>SD</i> = 75.91)	35.87 (<i>SD</i> = 28.40)
Total number of times each item was named across sessions	72.63 (<i>SD</i> = 19.33)	71.06 (<i>SD</i> = 29.84)
Grammatical class of the treated words	Nouns: 11; Verbs: 13; Mixture of nouns and verbs: 3	Nouns: 15; Verbs: 43; Mixture of nouns and verbs: 67
4. Variables describing the techniques		
Provision of semantic cue(s): Explicit	Yes: 23; No: 4	Yes: 64; No: 61
Provision of phonological cue(s): Explicit	Yes: 27	Yes: 94; No: 31

Provision of orthographic cue(s): Explicit	Yes: 27	Yes: 100; No: 25
Provision of phonological cue(s): Implicit	No: 27	No: 125
Provision of the target word in spoken form	Yes: 27	Yes: 121; No: 4
Provision of the target word in written form	Yes: 27	Yes: 80; No: 45
Spoken repetition of target word: Compulsory	Yes: 2; No: 25	Yes: 37; No: 88
Spoken repetition of target word: Contingent on response	Yes: 27	Yes: 109; No: 16
Copying of written target word: Compulsory	No: 27	Yes: ; No: 125
Copying of written target word: Contingent on response	No: 27	No: 125
Semantic cue(s) generation	No: 27	Yes: 8; No: 117
Phonological cue(s) generation	No: 27	Yes: 8; No: 117
Semantic task(s)	Yes: 22; No: 5	Yes: 10; No: 115
Orthographic task(s)	No: 27	No: 125
5. Variables describing the application of technique(s)		
Application of cues: Progression	Increasing: 25; Decreasing: 2	Increasing: 68; Decreasing: 26; Not applicable: 31
Application of cues: Contingent on response	Contingent on response: 5; Mixture of compulsory and response-contingent applications: 22	Contingent on response: 70; Compulsory: 11; Mixture of compulsory and response-contingent applications: 22; Not applicable: 22
Feedback on accuracy of the naming response	Implicit feedback: 27	Explicit feedback: 2; Implicit feedback: 100; No feedback: 23
Termination criteria	Fixed number of sessions): 27	Performance criterion: 1; Fixed number of sessions: 124
Mode in which actual therapy was delivered	Computer: 22; Person: 5	Computer: 12; Person: 113
Homework	No: 27	No: 125

Figure 2.2. Variables and their values of variable importance for improved spoken lexical-retrieval of treated items post-therapy in the short-term phase.

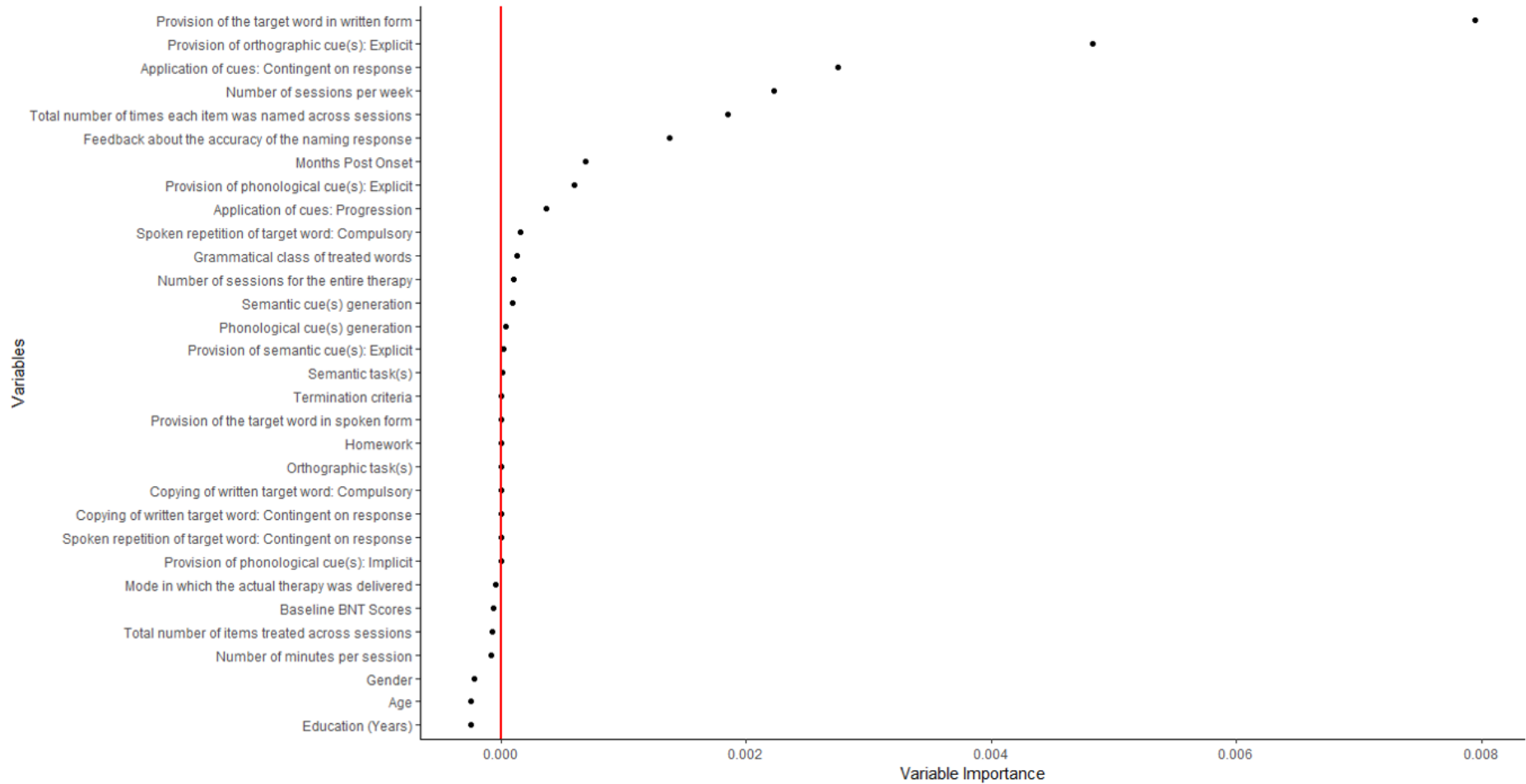


Figure 2.3. Variables and their values of variable importance for improved spoken lexical-retrieval of treated items post-therapy during the maintenance phase.

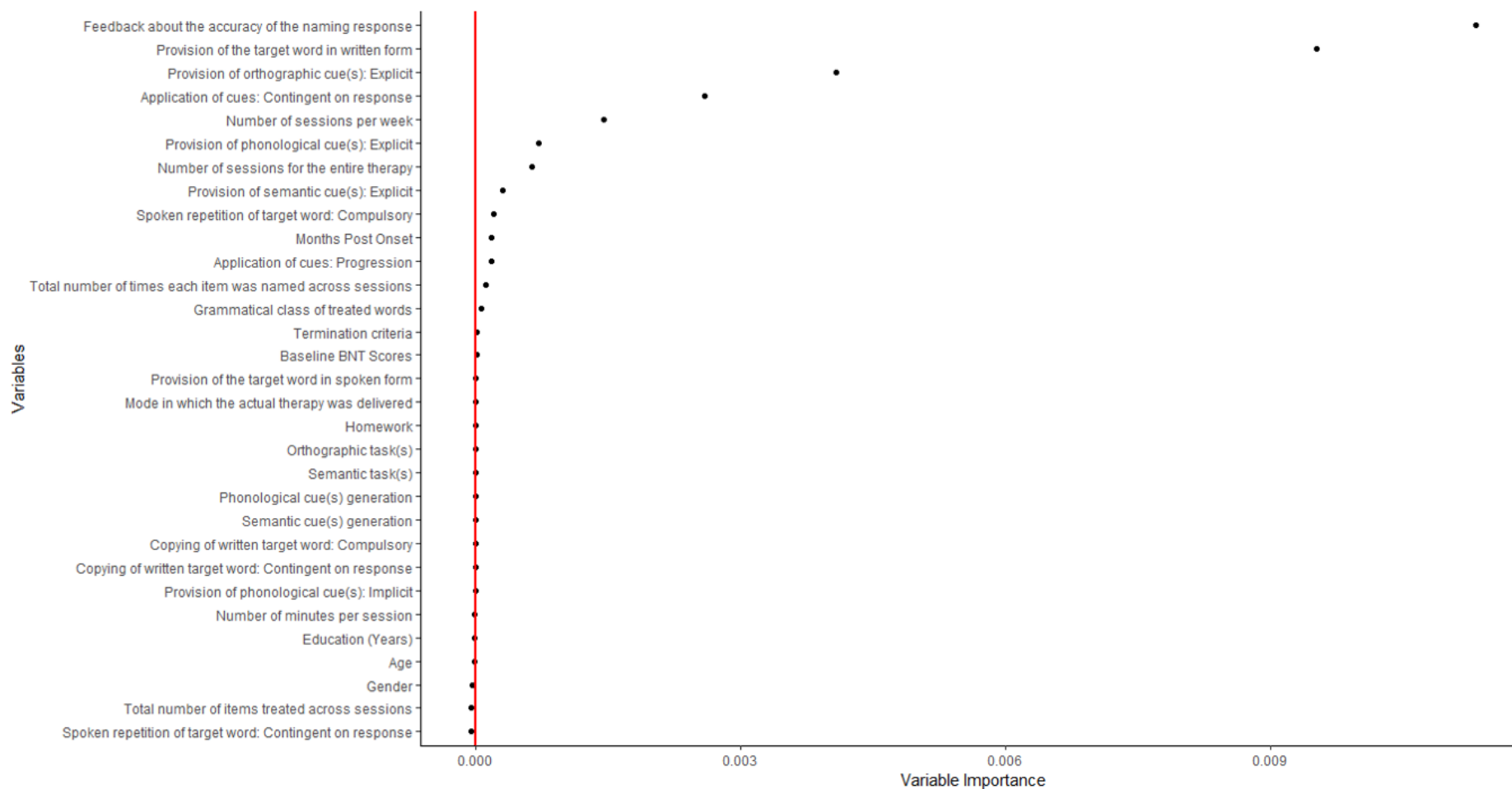
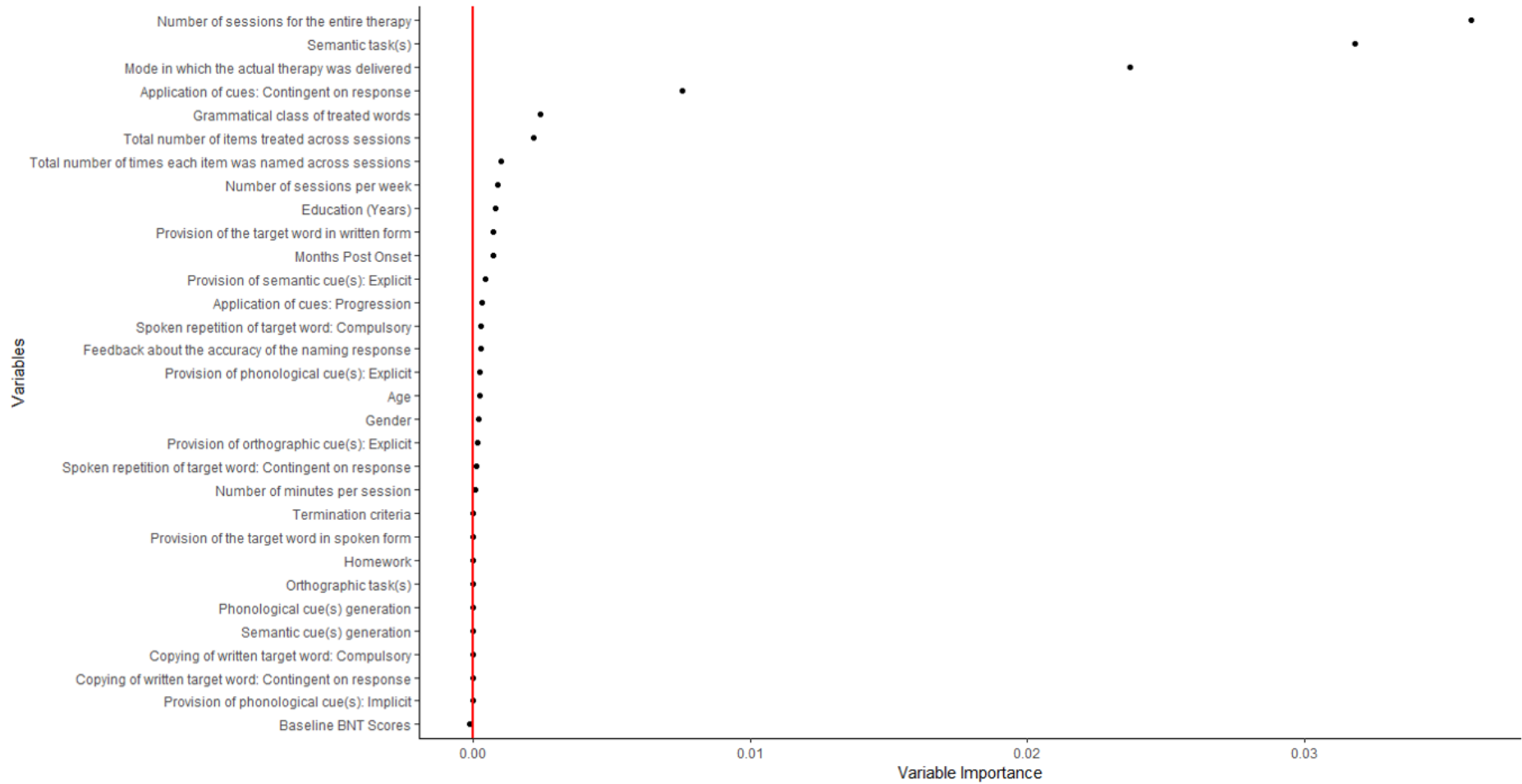


Figure 2.4. Variables and their values of variable importance for improved spoken lexical-retrieval of untreated items post-therapy in the short-term phase.



2.3.3 Successful treatment outcome of untreated items post-therapy in the short-term phase

There were 152 cases with published data for the individual's spoken naming of the untreated items three weeks after therapy. Out of these, 27 reported an improvement, while 125 saw no difference before and after therapy (see Table 2.8 for descriptive statistics). The representative random forest yielded the following diagnostic statistics [$C = 0.976$; OOB estimate = 0.086], which indicated that the selected random forest classification is also a robust predictive model. The correlation between the representative model's rankings and the overall median rankings is strong [$r_T = .987$, $p < .0001$]. The rankings according to 'variable importance' obtained for the representative random forest model are displayed in Figure 2.4. The variables – (1) Number of sessions for the entire therapy, (2) Use of semantic task(s), and (3) Mode in which the actual therapy was delivered (e.g., delivered on a computer or by a trained individual), emerged to be the top three predictors. Provision of target word in the written form was also amongst the top ten predictors.

2.4 Discussion

The multiple components making up a word-finding therapy are like the coloured tiles on a Rubik's cube. When the cube is twisted to coordinate colours on the top surface, all manners of unexpected transformations occur on the five remaining sides concealed from a fixed vantage point. This is why to crack the therapy riddle, there needs to be consideration of components on all sides. In this study, the multi-components in single-word spoken lexical-retrieval therapy were systematically reviewed and simultaneously accounted for using random forest analyses. The novel 'RITA' framework was proposed to guide a comprehensive coding of therapy components. In the process, evidence from 222 individual anomia cases were extracted from 32 studies. The final output were three sets of findings, one for each of the outcome measures, namely the improvement for treated items at the (1) short-term interim and (2) maintenance phase, as well as for (3) untreated items at the short-term interim.

In this paper, clinical assessment data were excluded as one of the predictors. As previously mentioned, this took into account the variety of tests employed in naming therapy studies, including bespoke assessments tailored to meet particular research needs (e.g., the unpublished 'Easy Naming Set' used by Mason et al., 2011). There are ongoing efforts to establish a core set of standardised tests to be used in aphasia research (e.g., Wallace et al., 2019). This move towards a more uniform battery of assessments is likely instrumental in driving future meta-analytic attempts forward.

Using published results as data points may result in inadvertent skewness due to publication bias. For instance, more cases of improvement than no improvement were reported for the treated items (Tables 2.6 and 2.7). However, two points are worth mentioning: Firstly, there is greater awareness by editors/reviewers to appreciate null results (see 'The importance of no evidence' [Editorial], 2019). In time, issues of positive publication bias would then be addressed. Secondly, the reverse is found for untreated items, where more cases of no improvement are published (Table 2.8). This serves to illustrate how the landscape of publication bias in aphasiology is slightly more complex, when the same papers are providing the data points for both treated and untreated items. This dichotomy therefore leaves open the possibility that what has been reported is indeed a genuine reflection.

The primary goal of this meta-analysis was to examine whether there were major therapy components that underlie treatment efficacy. There is a consistent theme across the three sets of findings: The role of orthography. Provision of the target word in the written form as well as the explicit provision of orthographic cue(s) were revealed to be particularly felicitous in the rehabilitation of treated items, be it in the short or longer term, as well as for untreated items in the short-term (Figures 2.2 to 2.4). It is also worth noting from the figures that the presentation of the written form itself is sufficient, without the ancillary need to copy the words out. Unfortunately, the orthographic approach per se is not well-explored within research¹³, with greater emphasis placed on semantic approaches (e.g., in Kiran & Johnson, 2008; van Hees et al., 2013) or phonological cueing and repetition (e.g., Biedermann & Nickels, 2008; Mason et al., 2011). In fact, in this review, none of the interventions published between January 2008 and May 2018, investigated the role of orthography exclusively. In saying so, the written form, had been frequently employed (and perhaps its clinical utility disguised) as part of a therapy using a mixture of cues (e.g., in Harnish et al., 2014, orthographic cue was used in step 2, but semantic and phonological part-word cues were applied in steps 5 and 6), or perhaps relegated to a supplementary role in the wider

¹³ To clarify, we are referring to the lack of studies that investigate the exclusive role of orthography, i.e., provide only the written form in treatment and is not accompanied by other cues (e.g., phonology). Orthographic cues are often explored in aphasia rehabilitation literature in the context of other techniques/cues, but it also means we cannot disambiguate the role of orthography from these accompaniments. We would also like to mention that the figures in Tables 2.6 to 2.8 do not contradict our statement on the rarity of research dedicated to the sole investigation of orthographic cues, compared to other approaches. In these tables, the numbers reflect the *number of participants*, not the number of studies. Snell and colleagues (2010), who provided both phonological and orthographic cues, was a case series of 13 participants. Rider and colleagues (2008), who used SFA, had three participants. This is not surprising because of the resource demands (e.g., time) necessitated when conducting SFA, which may explain the smaller sample sizes. Nevertheless, many participants were involved in research that examined written cues in the wider context of other techniques/cues. These high numbers further emphasise how few participants (in our case, zero participants across the 32 papers) undertook a therapy research that relied purely on orthography.

context of say, assisting verbal repetition during multimodal cueing (e.g., step 1 in Menke et al., 2009).

Further research will be necessary to examine why orthography works and how it operates within spoken naming therapy (Lorenz & Nickels, 2007). In the subsequent chapters, the role of the written form during spoken naming therapy will be clarified, first from the perspective of a facilitation trial (Chapters 3 and 4), before extending it into a full cueing therapy (Chapters 5 and 6). Facilitation involves limited applications of the cue, usually a one-off application, and the effect is measured not immediately but a while later (e.g., 15 minutes later). A therapy trial, however, will entail multiple applications of the cue akin to an actual intervention (see p. 820 in Howard, Patterson, Franklin, Orchard-Lisle, and Morton, 1985).

One useful finding to consider in clinical practice is that the participant's baseline naming performance did not emerge to be a strong predictor across the three outcome measures. Baseline naming performance here is indicated by the participant's Boston Naming Test score. This is potentially good news, because it suggests that participants across different levels of initial naming abilities can benefit from the various naming therapies included in the meta-analysis. In saying so, it is also possible that other participant factors are more important, like the individual's motivation or hardiness (Biel, Nitta, & Jackson, 2018), or perhaps the individual's cognitive reserve, like attention (Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010), or short-term memory/working memory (Wright & Fergadiotis, 2012). If a set of cognitive assessments were made standardised for aphasia research, it might be possible to include these factors for a meta-analysis in the future.

Another interesting observation relates to the clinical practice of giving clients homework. In all three outcome measures, the component 'homework' did not show substantial value. It thus appeared that in the papers synthesised, giving the participants homework did not actually have an impact on their eventual naming performance. This goes against the common clinical practice of setting clients extra self-practice outside therapy hours. The reasons why homework produced weak amount of 'variable importance' are not clear. It could indeed be that input during therapy suffices. Alternatively, perhaps having individuals with anomia complete homework without appropriate direct supervision by a trained person, does not add therapeutic value. An intervention study manipulating the administration, i.e., completing homework with and without supervised support versus the absence of homework, could help tease apart these accounts.

Many of the findings from this meta-analysis also corroborate what had been previously found. For instance, published evidence suggests that a language approach whose focus is on word forms (e.g., orthography or phonology) tended to result in more treatment-specific effects, whereas an approach that targets the semantics might result in generalization to untreated items (e.g., Nickels and Best, 1996a). This trend was replicated in this chapter, with the explicit provision of orthographic and phonological cues appearing in the top ten variables predicting successes for treated items in the short-term and maintenance phases (Figures 2.2 and 2.3; Tables 2.6 and 2.7), but only the orthographic format, specifically the provision of the entire target word in its written form, appearing within the top ten rankings for untreated items in the short-term (Figure 2.4). A different pattern was found for the use of say, semantic tasks. On the list of variables pertinent for untreated items, 'semantic task' secured the second highest imputed value for variable importance (Figure 2.4), but it did not appear to be a predictor for treated items. In saying so, the lack of generalization effects for phonological and orthographic approaches might also reflect a publication or research bias. This is because aphasia researchers tended not to control for orthographic or phonological similarities between treated and untreated items. Instead, one suspects that in most published research, greater shared semantic overlaps might be present between treated and untreated sets of items, although this maybe unintentional (e.g., see the stimuli lists appended in Wright, Marshall, Wilson, and Page, 2008). Still, there is tentative evidence that generalization at the word form level can occur (see Biederman, Blaken, & Nickels, 2002; Biederman and Nickels, 2008), but this warrants further investigation.

Another main finding replicated in this meta-analysis, is the importance of intensity, expressed in the form of the total number of times an item was trained. The 'total number of times an item was named across sessions', for instance, appeared to influence the naming outcome of treated and untreated items in the short term (Figures 2.2 and 2.4). This is also reported elsewhere (e.g., Harnish and colleagues, 2014, who showed that increasing the total number of 'teaching episodes' could improve naming for all their participants).

Lastly, in outlining possible 'active' therapy components making up a typical naming intervention, a novel 'RITA' framework was proposed: (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts (Tables 2 to 5). This structure is helpful for either reviewers evaluating interventions by breaking them down into their subparts or to guiding clinicians and researchers to think through the nuts and bolts in therapy design. By necessity, the list of therapy components compiled is not exhaustive, but clinically purposeful. The 'RITA' framework is supplied in Appendix 2.3, to facilitate interested users and to advance its use. The list of therapy components will always benefit

from constant refinement, but the framework represents a practical start. Although designed specifically with adult word-finding therapy in mind, the 'RITA' structure can easily be adapted for use in other areas within aphasia and dysarthria therapies, or extended to the paediatric population.

Addendum to Chapter 2: Short update on spoken word-finding intervention research between June 2018 and March 2022

The meta-analysis described above examined research between January 2009 to May 2018. To understand the changes and continuity in anomia therapy research published since then, a simplified search is conducted using one of the primary databases, PubMed.

Method

Information sources and search

The same search strategy used previously (Section 2.2.1, p. 34) was entered into the PubMed database, with the dates filtered between June 2018 to 18 March 2022; the latter being the date the search was conducted. PubMed was chosen as it is the official database maintained by the U.S. National Library of Medicine at their National Institutes of Health. It is known to be a comprehensive database, particularly in the areas of medicine and clinical science.

Study selection and eligibility criteria

The overall search generated 546 entries. These were then individually screened by the author using the eligibility criteria described earlier in Section 2.2.2 (p. 35). Out of these, three papers were eligible. The details of the screening stages are found in Figure 2.5.

Results

Although over 500 entries were generated by PubMed, most were ineligible for various reasons. These include how their therapies were not language-based per se (e.g., Cavanaugh, Quique, Dickey, Hula, Boss, and Evans, in press, involved teaching participants cognitive strategies like mental imageries to reduce their stress) or that the researchers did not measure their outcomes using picture naming. For the latter, an example is Renvall and Nickels (2019), whose therapy focused on repeating the spoken whole word in the presence of the picture, which fulfilled the inclusion criteria for language-based therapy, but their outcome was measured using generative naming – a verbal fluency task whose dynamics, as the authors themselves described, is different from direct picture naming (pp. 1366-1367). Consequently, only three studies remained eligible. Table 2.9 summarises their therapies and main findings.

Figure 2.5. PRISMA flow chart for the update on systematic review searched using PubMed (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

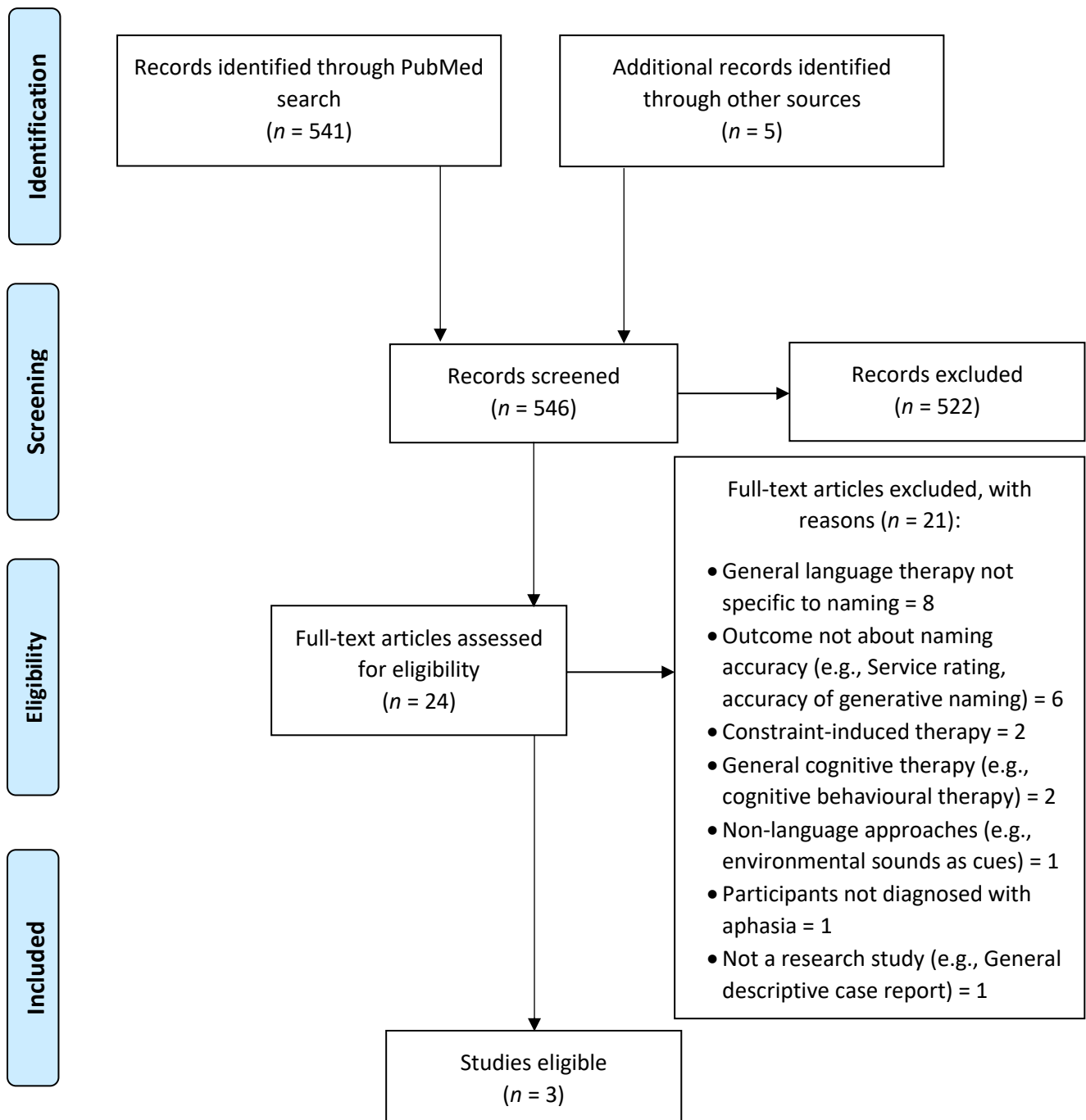


Table 2.9. Summary of spoken word-finding therapy studies involving semantic, phonological and orthographic language-based approaches published between June 2018 to mid-March 2022.

Study (Chronological)	Description of therapy	Main findings (post- vs. baseline) based on the picture naming paradigm
1a. Conroy, Drosopoulou, Humphreys, Halai, and Lambon Ralph (2018): Therapy 1 (called 'Standard Therapy' in the paper) <i>n</i> = 20 (11 males, 9 females; mean age = 65.2 years (<i>SD</i> = 11.7))	Therapy technique: Phonological and orthographic cueing – The target word is always provided in both spoken and written formats simultaneously at the end of the 10-seconds naming trial. If the response given during the trial was incorrect, additional spoken cues were provided subsequently and they progressed from initial consonant and vowel, followed by part-word (e.g., initial syllable), and then the whole word. The authors did not state the time allowed for naming before cues were applied. Treated language: English Treatment frequency: 6 sessions (2 sessions per week for 3 weeks). Session duration varies (30 to 50 minutes, p. 1818).	• Data were analysed at the group level. Individual level analyses were not conducted. • Significant improvement on the accuracy and response time taken to name treated items observed at 1 and 4 weeks post-therapy. • No significant changes in the accuracy of naming untreated control items. • Significant improvement on the response time taken to name untreated items across baseline to 4 weeks post-therapy. No related post hoc tests were reported. • However, 'Standard Therapy' resulted in significantly faster responses than untreated control condition.
1b. Conroy, Drosopoulou, Humphreys, Halai, and Lambon Ralph (2018): Therapy 2 (called 'Repeated, Increasingly-Speeded Production (RISP) Treatment' in the paper) <i>n</i> = 10 (subset of the 20 participants described above; characteristics of subset not provided)	Therapy technique: Phonological and orthographic cueing with shortened time limit – The target word is always provided in both spoken and written formats simultaneously at the end of the naming trial. Time allowed for naming shortened systematically across the sessions from 3 seconds initially to 1 second in the final session. If the response given during the trial was incorrect, participant was asked to repeat the whole word (spoken cue) after the computer or experimenter for three times (Table 1, p. 1819). Treated language: English Treatment frequency: 6 sessions (2 sessions per week for 3 weeks). Session duration varies (30 to 50 minutes, p. 1818).	• Data were analysed at the group level. Individual level analyses were not conducted. • Significant improvement on the accuracy and response time (RT) taken to name treated items observed at 1 and 4 weeks post-therapy. • No significant changes in the accuracy of naming untreated control items. • Significant improvement on the response time taken to name untreated items across baseline to 4 weeks post-therapy. No related post hoc tests were reported. • However, 'RISP Treatment' resulted in significantly faster responses than untreated control condition.
2. Lavoie, Bier, and Macoir (2019):	Therapy technique: Modified Semantic Feature Analysis i.e., (1) Answer 8 "Yes/No" semantic-verification questions related to the target noun. Questions were in the written form. (2)	• Significant improvement on the accuracy of naming treated items observed immediately within 1-week post-therapy for all 4 participants, and up till 2

$n = 4$ (P.R.: 73-year-old female, 17 months post-stroke; S.T.: 72-year-old male, 8 years post-stroke; L.M.: 59-year-old male, 17 months post-stroke; C.G.: 63-year-old man, 14 months post-stroke)	Attempt to name the picture. (3) Participants instructed to click an icon to play an audio clip of the target word for spoken repetition. This was independent of whether the naming attempt was accurate or not. Treated language: French Treatment frequency: 16 sessions (4 sessions per week for 4 weeks). Session duration varies (30 to 90 mins, pp. 257-258): Self-administered at own pace on smart tablet.	months post-therapy for 3 participants (P.R., ST., L.M.).¹⁴ No significant changes in the accuracy of naming “untreated but exposed” items and “untreated unseen” control items were observed for 2 participants (P.R., ST.), when tested immediately within 1-week and up till 2 months post-therapy.
3. Kristinsson et al. (2021): $n = 50$ (31 males, 19 females; mean age = 61.8 years ($SD = 11.4$))	Therapy technique: Combination of phonological therapies, i.e., (1) Phonological components analysis (identify phonological features, i.e., first sound, number of syllables, last sound, rhyming vowel, vowels in the first and last syllables); (2) Phonological production task (sorting pictures according to the number of syllables of the target words; identifying various phonological features using a pair of targeted nouns/verbs, e.g., first syllable-last syllable, last syllable-last syllable, first syllable-first sound etc.). Once features were identified, participants should blend the syllables/sounds together; (3) Phonological judgement task (judge whether pairs of words are matched on various phonological features e.g., same number of syllables, same initial syllables, rhyme etc.). Treated language: English Treatment frequency: 15 sessions (5 one-hour sessions per week for 3 weeks).	- Data analysed at the group level. Individual level analyses not conducted. - Data not separated into treated/untreated items. A standardised picture naming test (Philadelphia Naming Test, PNT) was administered instead, to track changes before and after therapy. Significant improvement on the accuracy of naming PNT items when tested immediately within 1-week post-therapy.¹⁵

¹⁴ In their paper, the authors claimed that “for all [four] participants, the treatment resulted in a significant improvement for both sets of trained words that was maintained 2 months after the end of treatment.” This is likely an overstatement, because for the last participant, C.G., both his performances on “untreated unseen” control items and “exposed but untreated” items were also significantly better post-therapy than at baseline. Post hoc analyses at two months post-therapy, also did not find any difference between performances on his treated items against either that of the “untreated unseen” or “exposed but untreated” items, suggesting that for C.G., there was no strong evidence for a treatment effect at the two-month post-therapy maintenance phase. The rest of the participants, however, appeared to maintain the treatment effects two months later, including L.M., whose performance of treated items remained significantly better than the “untreated unseen” controls as well as “exposed but untreated” items two months later (pp. 259-260).

¹⁵ Results here refer to the analyses using raw values reported in the Supplementary Results ($p = .008$), not the proportion of maximal gain (PMG) calculated using averaged baselines and reported elsewhere in their paper. This is because Kristinsson et al. (2021) mentioned the potential bias in their PMG values “inflating change scores for participants with mild aphasia” (p. 13). The second semantic therapy is excluded, because it involves elements of constraint-induced aphasia therapy and sentence production. Also, only the first post-therapy assessment time point is included because the second post-therapy assessment captures combined influences of phonological and semantic approaches. This is especially so given how the authors measured outcome using a general standardized naming test, which neither differentiated between treated/untreated items nor the therapy type.

Discussion

This search update is meant as brief review of anomia intervention published since the completion of the meta-analysis in Chapter 2. The three studies that fulfilled the eligibility criteria are a mix of small-*n* and larger-scale RCT designs. Kristinsson and colleagues (2021) belonged to one of the large-scale RCT studies (i.e., the “Predicting Outcome of Language Rehabilitation in Aphasia” (POLAR) project in this case) that just started releasing their data. Interestingly, and perhaps not surprisingly, one of the main takeaways from Kristinsson and colleagues (2021) was the heterogeneity in performance of people with anomia (Figure 5 on p. 10 in their paper). The researchers reported that although “between-group analysis revealed a superior effect of the semantic treatment on average, some participants showed a clear benefit of the phonological treatment (p. 7),” suggesting there are limitations to group findings – a point that will be revisited in subsequent chapters.

In addition, there appears to be a continued emphasis on semantic and phonological approaches. Again, none of the studies exclusively investigated the role of orthography – which is a running theme since 2009, when the meta-analysis begun. Conroy and colleagues (2018) did integrate orthographic cues within their two therapies, but given their design, it is not possible to distinguish the contribution of orthographic from phonological cues in their results.

Overall, it seems that the earlier call for studies to specifically investigate the role of orthography in spoken word-finding therapies remains relevant. This is a challenge taken up by the subsequent chapters.

Chapter 3

Comparing the facilitation effects of orthographic cues with or without the phonological components – Group-level participant analyses

3.1 Introduction

One of the surprising revelations in the previous chapter is the usefulness of orthographic cues. This meta-analytic finding is corroborated by recent findings by Brogan, Godecke and Ciccone (2020), who analysed live video recordings of hospital therapy sessions. They found that semantic cues were relatively common (on average, applied 12.9 times per session), yielding a success rate of 13.2%. In contrast, orthographic cues were employed much less frequently (on average, applied 4.1 times per session), yielding a success rate of 43.9%. After adjusting for the respective cue's average number of applications per session, orthographic cues (adjusted proportionate rate of success = 2.08) were in fact as effective as semantic cues (adjusted proportionate rate of success = 1.92), yet there is scant information on orthographic cues. Together, these findings suggested a need for further research into the optimal use of orthographic cueing in clinical contexts. Thus, the broad aim of the next two chapters is to delve deep into the outcomes of various cues with an orthographic element by first, tracking the facilitation effects of different combinations of these cues with a written component across different time points, and exploring the language skills associated with orthographic cueing during facilitation.

Chapters 3 and 4 are based on a facilitation study involving different compositions of orthographic cues. As mentioned in Chapter 2, facilitation is defined as an investigation involving limited cue applications, typically a one-off application. Readers can refer to Howard, Patterson, Franklin, Orchard-Lisle, and Morton (1985) for a discussion on the importance of pared-down facilitation studies to draw more definitive conclusions. Briefly, some advantages include how their minimalist setup enable effects from a single trial of a specific cue to be isolated, which is instrumental for developing understanding. Their smaller-scale also allows for techniques to be pilot-tested, before full therapy trials lasting several weeks per participant are launched.

Table 3.1 summarises the slim body of facilitation studies available on orthographic cueing, which involve participants with anomia. Aside from two instances of negative results reported by Bruce and Howard (1988), there is otherwise general agreement that orthographic cues can significantly benefit one's spoken word retrieval of treated items.

Table 3.1. Summary of facilitation studies involving orthographic stimuli only and with spoken output. All studies focused on spoken picture naming accuracy as the primary outcome.

Study (Chronological)	Description of facilitation study	Main findings (post-facilitation vs. final baseline, unless specified otherwise*) based on the picture naming paradigm
Bruce and Howard (1988): Referred to as 'Letter pointing' condition in the paper <i>n</i> = 20 participants with Broca's aphasia post-stroke	Facilitation technique: Orthographic task – A letter board with 9 letter choices (written) was provided, if the participant was unsuccessful in naming the picture within 5 seconds Number of cue applications provided per item: 1 Unit: Initial letter Treated language: English	*Baseline and post-facilitation outcomes were not compared in the paper. Instead, the authors compared facilitation outcomes against the outcomes from 'extra time (with no cues)' provided. When assessed within the 5 seconds from which facilitation condition or 'extra time' was provided: • For 7 participants: No significant difference found on treated items between 'letter board' and 'extra time' conditions. • For 2 participants (PAC, CHA): Significant decrease in accuracy on treated items for 'letter board' vs. 'extra time'. • For 11 participants: Comparisons between 'letter board' and 'extra time' were not provided, as the focus of their findings were on a significant difference between a third condition 'initial phoneme' (outside the scope of this paper) and 'extra time'.
Howard and Harding (1998): Referred to as 'Condition (ii): Use of alphabet board' in the paper <i>n</i> = 1 (SD: 46-year-old female, 20 weeks post-stroke; site(s) of brain lesion(s) not provided as brain scans were not available)	Facilitation technique: Orthographic task – An alphabet board was provided (number of letter options not stated in the paper), if the participant was unsuccessful in naming the picture within 5 seconds Number of cue applications provided per item: 1 Unit: Initial letter Treated language: English	*Baseline and post-facilitation outcomes were not compared in the paper. Instead, the authors compared facilitation outcome against the outcome from 'extra time (with no cues)' provided. When assessed within the 5 seconds from which facilitation condition or 'extra time' was provided ('Set A' analysis): • SD's naming on treated items with 'alphabet board' was significantly better than 'extra time'.
Howard and Harding (1998): Referred to as 'Condition (iii): Written Letter' in the paper <i>n</i> = 1 (SD: 46-year-old female, 20 weeks post-stroke; site(s) of brain lesion(s) not provided as brain scans were not available)	Facilitation technique: Orthographic cueing – Written 'initial letter' cue provided if the participant failed to name the picture correctly within 5 seconds Number of cue applications provided per item: 1 Unit: Initial letter Treated language: English	*Baseline and post-facilitation outcomes were not compared in the paper. Instead, the authors compared facilitation outcome against the outcome from 'extra time (with no cues)' provided. When assessed within the 5 seconds from which facilitation condition or 'extra time' was provided (Set B analysis): • SD's naming on treated items with 'written initial letter' cue was significantly better than 'extra time'.

Avila, Lambon Ralph, Parcet, Geffner, and Gonzalez-Darder (2001): <i>n</i> =1 (FR, 31-year-old man, Spanish/Catalan speaker, with anomia secondary to temporal lobe seizures and a cavernoma in the left superior temporal lobe)	Facilitation technique: Pre-stimulation – Written whole word presented for 33.4 ms, followed by 3 sets of masks (e.g., @@@@ @@@) lasting 53.1 ms each, before the picture was presented for naming. Number of cue applications provided per item: 1 Unit: Whole word Treated language: Spanish	• Significant increase in the number of items correctly named following facilitation (priming) compared to unprimed trials, when assessed 5 seconds post-facilitation. • Unprimed trials refer to trials in which XXXXXXXX replaced the written whole word during pre-stimulation. • Items were divided into sets 1 and 2 for naming. In session 1, set 1 received facilitation and set 2 did not (unprimed). In session 2, it was the reverse. So, each item was named once with facilitation and once without facilitation. The two sessions were a year apart and session 2 was conducted post-surgery.
Best et al. (2002): Referred to as 'Written CV' in the paper (Condition 4a, p. 155) <i>n</i> = 11 Participants were over one-year post-stroke with anomia	Facilitation technique: Orthographic cueing – Written CV cue provided if the participant failed to name the picture correctly within 5 seconds Number of cue applications provided per item: 1 Unit: CV Treated language: English	*Baseline and post-facilitation outcomes were not compared in the paper. Instead, the authors compared facilitation outcome against the outcome from 'extra time (with no cues)' provided. At the group-level (only group-level analyses reported): • Significant improvement on treated items when assessed within 5 seconds after the cue was administered (labelled as 'immediate' in the paper) and 10 minutes post-facilitation (labelled as 'delayed' in the paper) • No difference between single CV cue vs. choice of two CV cues when assessed 10 minutes post-facilitation (i.e., no choice effect across all types of cues, including Conditions 4a and 4b, see column on left in this table)
Best et al. (2002): Referred to as 'Written CV' in the paper (Condition 4b, p. 155) <i>n</i> = 11 Participants were over one-year post-stroke with anomia	Facilitation technique: Orthographic task – A choice of two written CV cues presented if the participant failed to name the picture correctly within 5 seconds Number of cue applications provided per item: 1 Unit: CV Treated language: English	*Baseline and post-facilitation outcomes were not compared in the paper. Instead, the authors compared facilitation outcome against the outcome from 'extra time (with no cues)' provided. At the group-level (only group-level analyses reported): • Significant improvement on treated items when assessed within 5 seconds after the cue was administered (labelled as 'immediate' in the paper) and 10 minutes post-facilitation (labelled as 'delayed' in the paper) • No difference between single CV cue vs. choice of two CV cues when assessed 10 minutes post-facilitation (i.e., no choice effect across all types of cues, including Conditions 4a and 4b, see column on left in this table)

Lorenz and Nickels (2007):

$n = 3$

Participants (MCB, KCC, JUE) were English monolinguals. Each had a left-sided stroke, at least 3.5 years post-onset.

Facilitation technique: Pre-stimulation with written initial letters
i.e., The initial letter was printed on screen for 600 ms to be read, before it was removed. The picture was then provided for naming.

Number of cue applications provided per item: 3

Unit: Initial letter

Treated language: English

Facilitation was divided into Phases 1 and 2 (three sessions in each phase).

Only results from Phase 1 were reported here as it was the only phase that involved all three participants.

Conclusions drawn from Phase 2 were also generally consistent with what was found in Phase 1.

The authors also manipulated phonological transparency of the initial letter cues and found it has no effect (described within text later).

For MCB and KCC:

- **Significant improvement on treated items** when assessed **directly after cue was administered** (labelled as 'immediate' in the paper) and **15-20 minutes post-facilitation** (labelled as 'delayed' or 'd' in the paper) for **each of the three sessions**, as well as **one day after completing three sessions** (labelled as 'delayed 24' or 'd24' in the paper).
- **Significant improvement on treated items** when assessed **one week later for MCB, but not KCC**.

For JUE:

- **Significant improvement on treated items** when assessed **directly after cue was administered** (labelled as 'immediate' in the paper) and **15-20 minutes post-facilitation** (labelled as 'delayed' or 'd' in the paper) for **sessions 2 and 3** (all results were not significant on session 1), as well as **one day after completing three sessions** (labelled as 'delayed 24' or 'd24' in the paper).
- JUE was not assessed one week after receiving the facilitation sessions.

For MCB, KCC and JUE:

- **No significant changes** in performance on all **untreated items** (unseen controls) when assessed **one day or one week after completing three sessions** (one day is labelled as 'delayed 24' or 'd24', and one week as 'BL_A1' in the paper). Only MCB and KCC were tested one week after their three facilitation sessions.
-

However, there appears to be a lack of consistency in terms of implementation. For instance, different studies operationalised the use of orthography stimuli variously. Some adopted a task-based approach, like getting the participant to choose the correct letter from a range of letter choices in the form of alphabet boards (e.g., Bruce & Howard, 1988), and others relied on orthographic cueing (e.g., Lorenz & Nickels, 2007). Best and colleagues (2002; Table 3.1) included both approaches in their study. Usefully, the authors compared the accuracy outcomes from 'single written CV cue' (Condition 4a in Best et al., 2002) against 'a choice of two written CV cues' (Condition 4b), and found that single written CV cueing was sufficient to achieve significant naming benefits. They showed that effects for the two conditions were statistically comparable, though qualitatively, performances on the 'single written CV cue' condition surpassed that of 'a choice of two written CV cues' (see Figure 1 on p. 156 in Best et al.). In view of their findings, the facilitation approach used in the next two chapters of this thesis will rely on a single cue. This is also in keeping with the use of a pared-down experimental approach for sake of clearer conclusions, as advocated by Howard and colleagues (1985) mentioned earlier.

However, it appears that there is still no consensus on what constitutes an optimal unit in orthographic cueing. Different studies in Table 3.1 employed different units of written cues, ranging from single letters (e.g., Condition (iii) in Howard and Harding, 1998) to the entire word (e.g., Avila et al., 2001). Therefore, it will be helpful to determine the unit size of a successful orthographic cue.

Another open question pertains to the duration of the facilitation effect. Understanding how long the facilitation lasts can help therapists plan a more time-efficient and realistic treatment schedule. Interestingly, not all studies investigated outcomes beyond the immediate few seconds after the facilitation cue was administered (e.g., Avila et al., 2001; Howard & Harding, 1998; see Table 3.1). Only one study involving a single application of cue examined outcomes at a slightly delayed period, i.e., about 10 to 15 minutes post-facilitation, and results showed that naming improvement was maintained (Best et al., 2002). However, it remains unclear if the effects can extend beyond the same day.

Finally, understanding how a specific cue operates can help determine when it may be appropriate for an individual. This includes asking practical questions, like: Does the client possess the requisite language abilities to profit from that written cue? Which linguistic skills ought to be prioritised for testing to establish the client's responsiveness towards a targeted cue?

To shed light on the underlying language skills, Best and colleagues (2002) correlated their participants' performances on baseline tasks with facilitation outcomes and found a significant association between grapheme-phoneme-conversion (scored on the number of initial phonemes of nonwords correctly read) and the ability to benefit from written cues. Thus, they inferred that orthographic cueing operated by the sub-lexical route. In contrast, Lorenz and Nickels (2007; also see Table 3.1 for other details) manipulated the phonological transparency of the initial letter cues. Their crucial contribution was to demonstrate how initial letters with non-transparent phonological mappings (e.g., 'k' for 'knight') were equally effective as cues when compared to letters with transparent mappings (e.g., 'd' for 'dog'). Unlike Best and colleagues' sub-lexical account, they showed that individuals with impaired sub-lexical nonword reading, could still benefit from initial letter cues through the lexical-semantic pathway.

Evidence for a third pathway was also provided by Howard and Harding (1998). Their participant's, SD's, picture naming improved when cued using either the initial written letter or with an alphabet board (Table 3.1). The authors demonstrated that the physical written form was necessary, because SD did not benefit when cued using the spoken letter name. As SD already had excellent semantic retrieval through pictures, the authors surmised it was unlikely the additional benefits from the orthographic cue trickled through the lexical-semantic route. Also, because SD had very poor grasp of grapheme-phoneme correspondence (she scored zero on letter sounding), the sub-lexical route proposed by Best and colleagues (2002) was also ruled out. Instead, the authors argued that the most plausible account is the exploitation of a third route – the direct lexical pathway linking visual input lexicon with the phonological output lexicon. By combining partial information from the written letter cue through this third pathway with the partial information drawn from the intact semantic system, the target word is made available in the participant's spoken output during picture-naming. A detailed discussion of these three pathways in relation to a language model is presented in the next chapter (paths 'A' to 'C' on Figure 4.1, p. 99).

3.1.1 Three primary research questions and an overall hypothesis

Four (three primary, one secondary) research questions were thus developed, based on the review above.

Primary research question (1): At the participant group level, which unit of the orthographic cue (initial letter versus whole word) will result in significant improvement on the three main naming outcomes (initial/final accuracies, RT)?

As shown in Table 3.1, there is no consensus, with evidence supporting the use of single letters (Howard and Harding, 1998), CV syllables (Best et al., 2002) and whole words (Avila et al., 2001). Existing clinical practice does not favour or discriminate between any orthographic unit. To address this research question, the two extreme sizes on the orthographic unit spectrum, initial letter versus the whole word, were contrasted. A successful cue is operationalised as the cue condition that is able to deliver improved outcomes on a post-facilitation time point relative to baseline. This means ideally, it should result in faster (RT) and/or more accurate responses, relative to baseline. Because question (1) is an open empirical question, whichever is the stronger unit should therefore achieve higher initial and final accuracies, as well as faster response times.

Primary research question (2): At the participant group level, which mode of cueing (orthography only versus multiple modalities with orthography and phonology combined) will result in significant improvement on the three main naming outcomes (initial/final accuracies, RT)?

At present, practising therapists are encouraged to use multi-modal cueing during naming therapy (pp. 412-414 in Coelho, Sinotte, & Duffy, 2008). This includes applying orthographic (e.g., the written form 'cat') and phonological cues (e.g., /kæt/) simultaneously to maximise rehabilitation gains. To the best of this author's knowledge, there is no direct evidence supporting this clinical practice. People with aphasia often have other cognitive comorbidities (e.g., attention deficits). Covering multiple modalities may potentially overwhelm and/or distract the cognitively-impaired client. Indirect evidence for multimodal cues in naming arrive in the form of studies, like Snell and colleagues (2010), who simultaneously applied written and phonological cues (e.g., 'b' with /bə/) to investigate other research questions (e.g., large versus small number of treated items). There is yet to be an explicit evaluation of written cues against combined written and phonological cues.

In response to this research question, the number of modalities - 'orthography only' versus 'orthography and phonology in combination', was manipulated. Crossing the two factors, i.e., unit of the orthographic cue (Research question 1) against modality (Research question 2), would generate four types of facilitation cues (see cue conditions under Section 3.2.4.3, p. 73). Based on prevailing clinical practice favouring multi-modal cues, 'orthography and phonology' in combination should produce superior outcomes (higher accuracies and/or faster RT), be it at the single letter or whole word level.

Primary research question (3): How durable are the facilitation effects of each combination of cue (unit x modality) across the participants?

Having empirical data on the effect's duration will help clinicians develop realistic treatment schedules. As discussed above, none of the studies involving a single application of cue investigated beyond 15-20 minutes post-facilitation (Table 3.1). It is clinically relevant to ask if the effects could extend beyond the day facilitation took place. To answer question (3), the participants were assessed twice; first at about 15-minutes after a single exposure to the facilitation cue and then, a week later.

Thus, for research questions (1) to (3), the overall hypothesis is: Assuming current clinical practice is correct, 'orthography and phonology' multi-modal cue combination should produce superior outcomes (higher accuracies and/or faster RT) *at both assessment time points (15 minutes and one week later)*, independent of the unit size of the cue.

3.1.2 A secondary research question and its hypothesis

Secondary research question (4): At the participant group level, what is the relationship between baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?

Lastly, to explore the residual language skills involved, correlations will be run between the baseline assessments and the facilitation gains (see below) from various orthographic cue combinations. The hypothesis is that participants should vary in the amount they benefit (if at all) from each cue condition, according to their language strengths/weaknesses. It is likely that individuals impaired on written language processing (e.g., difficulty reading printed words) may not benefit from the 'Orthography only' conditions. However, individuals with relatively intact spoken input language processing (e.g., intact auditory lexical decision) would be expected to profit from the 'Orthography and Phonology' cue conditions. Due to the correlational design in the context of a small sample size (details below), this analysis will be exploratory.

3.2. Method

3.2.1 Design

A case series approach ($n = 5$) with multiple baselines and control task design was adopted. The facilitation cue conditions followed a 2 (Unit of cue: 'Initial Letter' versus

'Whole Word') x 2 (Modality: 'Orthography' versus 'Orthography and Phonology') design, and were assessed twice post-facilitation (15 minutes versus 1 week later). Picture naming was chosen as the primary paradigm, for its common use in related research and within clinics. Nonword repetition was selected as the control task (PALPA 36; Kay, Colheart, & Lesser, 1992). This is because repeating nonsense syllables, should in theory, not be influenced by either orthographic and/or phonological facilitation cueing.

3.2.2 Participants

The inclusion criteria for a participant were as follow. The individual should: (1) be an adult (at least 18-year-old), (2) be a native speaker of British English, (3) have had a stroke at least six-months prior to the commencement of the study, (4) have difficulties with spoken word-finding after the stroke, and (5) have adequate auditory and visual acuity to interact with a laptop computer. Individuals with moderate or severe apraxia of speech, and/or diagnosed with neurodegenerative diseases (e.g., dementia) were excluded. Apraxia of speech was determined using the diadochokinetic task (based on Icht and Ben-David's, 2014, norms) and oromotor examination administered in-person by the author (a qualified speech and language therapist) during the initial session, as well as clinical records from the university's communication clinic.

Ethics was submitted and cleared at the university level (Project ID: 14189/001), before recruitment took place. Participants were recruited through the university's communication clinic register (see Appendix 3.1 for the recruitment poster sent out). Thirty individuals were contacted and twenty-five responded. Two individuals did not meet the criteria as they were impaired on speech but not language; seventeen declined participation due to conflicts in schedule or were simply not interested. Six were eventually recruited, but one participant's data could not be analysed as he failed to follow the instructions¹⁶. The final sample was made up of five participants (3 males, 2 females; age ranged from 55 to 69) and all gave informed consent to participate. Their background information is published in Table 3.2. Their sites of lesions were primarily in the region of the left middle cerebral artery (MCA). P5's lesion was in particular, large and encompassed the entire left MCA cortical territory, extending to the basal ganglia.

¹⁶ Participants were instructed that there was no need for additional practice beyond the facilitation session conducted. One participant, however, was conscientiously writing down the target words in his notebook after the facilitation session for home practice. This was discovered by the researcher on the last day during debriefing. The participant's data was therefore set aside and not coded for analyses.

Table 3.2. Participant information.

	Age	Gender	Ethnicity	Highest level of education	Lesion site(s)	Time post-onset (Years; months)	WAB score and diagnosis from the communication clinic registry ¹⁷
P1	65	M	Caucasian	Secondary School	Left MCA (lateral and inferior prefrontal cortex, with likely anterior temporal involvement)	8;3	46; Broca's aphasia
P2	55	F	African descent	Secondary School (Third Year)	Left cerebral vascular accident (CVA; no further information on lesion location available)	4;1	75.5; Anomia
P3	69	F	Caucasian	University	Left total anterior circulation infarct	4;1	53.9; Broca's aphasia
P4	58	M	Caucasian	University	Left MCA (no further information on lesion location available)	4;7	51.9; Broca's aphasia
P5	69	M	Caucasian	Secondary School	Left MCA (large lesion involving the entire left MCA cortical territory, underlying white matter and basal ganglia)	7;1	Not available; Broca's aphasia

¹⁷ With the exception of WAB Aphasia Quotient (AQ), scores on all other baseline assessments in Chapters 3 and 4 were collected from participants just prior to the facilitation session. The WAB AQ scores reported in the facilitation study are extracted from the university's communication clinic registry. This is perhaps why P1's AQ seemed low, yet his naming performance during actual baseline (Chapter 4) was stronger than others who had higher AQ scores (e.g., P3). Afterall, it is possible that P1 had simply improved since the time he was assessed on the WAB. However, due to data confidentiality, all data were de-identified after data collection, there is no means of re-tracing P1's name and clarify when his WAB information was collected.

All five participants were community-ambulant. According to their family members, none of them suffered from significant visual or hearing impairments, and that hearing aids were not required. These were further verified in the subsequent sessions: All five participants scored full marks on letter discrimination task in ‘Psycholinguistic Assessment of Language Processing in Aphasia’ subtest 18 (PALPA 18: Mirror Reversal; Kay et al., 1992) administered at baseline, and all of them could accurately repeat the spoken words played in the audio clips during preliminary sound checks (3/3) prior to practice trials in the facilitation naming experiment (full experiment described later). All five participants were also able to follow at least one-step simple instructions independently (6 out of 6 items) during oromotor assessment (Appendix 3.2). Combined, these suggested that each participant’s auditory perception was likely intact. Oromotor assessment also did not detect orofacial abnormalities (e.g., lingual weakness affecting articulation) in any participants. None had moderate or severe apraxia of speech.

3.2.3 Materials

296 stimuli nouns with coloured photographs were assembled from various normative studies (see Appendix 3.3 for a breakdown of sources). Only pictures with at least 80% name agreement were selected for use. However, because 42 pictures were sourced directly from royalty-free websites, they came without name-agreement norms. For these, name-agreement was established by having an independent sample of healthy native-English adults (4 males, 1 female; mean age = 52 ($SD = 7.97$)) name those pictures. Of the 42 pictures, 30 had identical names independently provided by at least 4 individuals, and these were included for subsequent use.¹⁸

In addition, only pictures with target words with initial phonemes represented by single written letters were chosen. This meant that pictures with target words whose initial sounds (e.g., /ʃ/) mapped onto double letters (e.g., ‘sh’) were omitted. The pictures were then digitized to a 500x500 pixels format with 24-bits depth for uniformity and maximum clarity, using GIMP 2.10.6 (The GIMP Team, 2019).

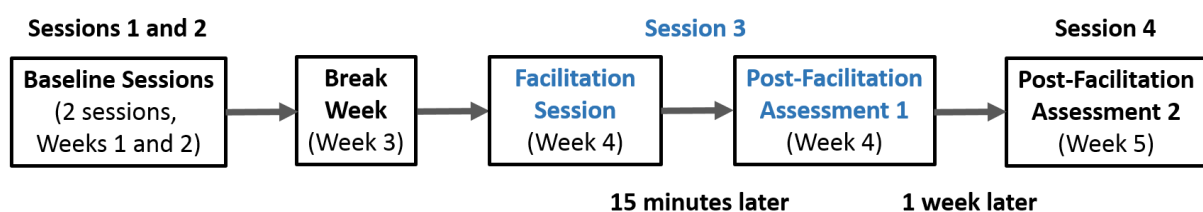
¹⁸ In an ideal situation, all pictures would be normed using the same individuals (say $n \geq 30$), but that is not always logistically feasible, e.g., manpower and timeline. Measures were taken to reduce potential issues/confounds (if any): (1) All photographs feature concrete nouns. This would reduce likelihood of ambiguity in responses, unlike say for abstract nouns or verbs. (2) Aside from the small subset of 42 pictures sourced from royalty-free websites and normed using a small sample (as this is *not* a norming study), the bulk of the pictures (254 pictures) used here were from published sources (Appendix 3.3) with norms collected from respectable sample sizes (smallest being $n = 39$ in Brodeur et al., 2010, to $n = 159$ in Adlington et al, 2009). (3) All norms, including the published sources, were collected from healthy English first-language speakers. (4) Importantly, alternatives were accepted as valid (e.g., ‘brolly’ for ‘umbrella’), as described in section 3.2.4.5.

For each target word, separate audio recordings were made for its initial phoneme and the word in its entirety. An English female native speaker read the phonemes and words aloud. Recordings were made in a sound-proof studio with a Rode NT-1A microphone at a sampling rate of 4,4100 Hz, at stereo and 16-bit setting for optimal speech sound recording.

3.2.4 Procedure

The entire study took five weeks for each participant. Participants were seen individually face-to-face in-person once per week for four sessions (see Figure 3.1 for the timeline). There was no session on the third week. Each session generally lasted around two hours with ample breaks in between tasks to reduce fatigue. There was a compulsory 15-minute break for refreshment and washroom around the one-hour mark. Additional breaks were determined by the participants: They were asked if a break was needed after each assessment/task was completed. They were also informed at the start of every session that they were allowed to take breaks during the session. The final session was shorter than the rest and lasted about an hour (see below for the tasks/assessments in each session), but as far as I can see, this did not impact on their performances. None of the participants complained of exhaustion either during the study. All sessions were conducted in a quiet room with minimal distractions.

Figure 3.1. Flowchart of the facilitation study.



3.2.4.1 Baseline assessments (Weeks 1 and 2)

In these two sessions conducted before facilitation cueing (Figure 3.1), standardised language assessments (including the control task) and spoken picture naming tasks were administered.

For the full list of standardised assessments, please refer to Appendix 3.2 that includes a short description of each subtest (or see Table 4.1 for a listing). The language assessments were primarily drawn from the Comprehensive Aphasia Test (CAT; Swinburn, Porter & Howard, 2004) and PALPA (Kay et al., 1992) language batteries. Both batteries were selected as these were established tools frequently used by speech-and-language therapists in UK, and that their underlying test design/construct were grounded in an empirically-verified psycholinguistic model (Kay, Lesser & Colheart, 1996).

In tandem with the facilitation study's aim to compare effects from orthographic and phonological cues, specific subtests were chosen such that they target the processing of written and spoken input stimuli at different stages (refer to the headings in Table 4.1 for the primary domain each subtest assesses). Briefly, these included assessments on the auditory processing of phonemes (e.g., 'PALPA 23: Spoken letter sound-written letter matching') and words (e.g., 'PALPA 5: Auditory lexical decision task'), assessments on processing of written letters (e.g., 'PALPA 18: Letter discrimination') and words (e.g., 'PALPA 25: Visual lexical decision task'), sub-lexical conversion of letters into phonemes either individually (e.g., 'PALPA 22: Letter sounding') or as a letter string ('PALPA 36: Reading nonwords'), as well as assessments on semantic processing (e.g., 'CAT 2: Semantic memory', 'CAT 8: Comprehension of written words').

To establish baseline naming performance, spoken naming for the 296 photographs was also collected twice prior to facilitation (referred to as baselines 1 and 2) using the picture naming paradigm, as depicted in Figure 3.2a below. To acquaint the participants with the procedure, there were also five practice trials (target words not featured) at the start of each naming trial. No cues or feedback were provided.

3.2.4.2 Post-facilitation assessments (Weeks 4 and 5)

Facilitation cueing took place on Week 4 (the process of facilitation is elaborated below in sections 3.2.4.3 and 3.2.4.4). Directly after facilitation, participants watched a short video of comic events without dialogue (the entire interim including video-watching took roughly 15-minutes). After watching, they received their first post-facilitation assessment and returned one week later (Week 5), for their second post-facilitation assessment.

In each post-facilitation session, picture naming trials for all items selected for the six conditions were administered without cues/feedback. On the participants' last session at Week 5, nonword repetition control task was also re-administered and they were debriefed.

3.2.4.3 Facilitation session (Week 4)

Six types of conditions were studied, which were:

- Condition (1): Initial Letter Cue: Orthography (e.g., 'a')
- Condition (2): Whole Word Cue: Orthography (e.g., 'apple')
- Condition (3): Initial Letter Cue: Orthography and Phonology (e.g., 'a' + /æ/)
- Condition (4): Whole Word Cue: Orthography and Phonology (e.g., 'apple' + /æpəl/)
- Condition (5): Repeated Exposure (i.e., Items not cued but encountered during facilitation), and
- Condition (6): Unseen Control (i.e., Items only encountered during assessments but not during facilitation)

Conditions (1) and (2) were 'orthography only' facilitation cue conditions, while conditions (3) and (4) represented multimodal facilitation cue conditions. The remaining two were meant as control conditions. Condition (5) captured improvements not related to the cue manipulation, and is thus important. This is crucial in the context of adult aphasia rehabilitation, as improvements can arise due to repeated encounters with the same items without treatment; this was first documented as a single case study based on JAW by Nickels (2002a) and observed subsequently in other individuals (e.g., Creet, Morris, Howard, & Nickels, 2019). To demonstrate that facilitation (or treatment in Chapters 5 and 6) was successful, effects from cueing conditions should be identified over and beyond the effects from repeated exposure for matched items. Condition (6), on the other hand, contained items encountered during assessments, but not during the facilitation process. Consequently, Condition (6) should capture potential generalisation effects (if any) to the untreated items.

For each participant, bespoke target words were chosen based on his/her baseline naming accuracy performance. Six sets of matched target words were created and randomly assigned into six conditions using Research Randomiser (Urbaniak & Plous, 2013). These six bespoke sets were also matched on the following outstanding variables, to the best of the author's abilities: (a) the individual's mean baseline performance in terms final naming accuracies, (b) lexical frequency (indexed by subtitle contextual diversity; Adelman, Brown, & Quesada, 2006; Brysbaert & New, 2009), (c) number of letters, (d) number of phonemes, (e) number of syllables, and (f) consistency (Balota et al., 2007). Separate one-way within ANOVA was conducted on each of the six variables; their results confirmed that the six conditions were sufficiently matched [$p > .05$ across conditions for each variable].

To keep participants motivated, an additional subset of words (about 20% of the cued items) that the individual managed to successfully name consistently across both baselines, were included as 'items for encouragement'. This led to the following item makeup: 36 items

(30 were treated items) per condition for P1, P4 and P5 respectively, 19 items (16 treated items) per condition for P2, and 30 items (25 treated items) per condition for P3.

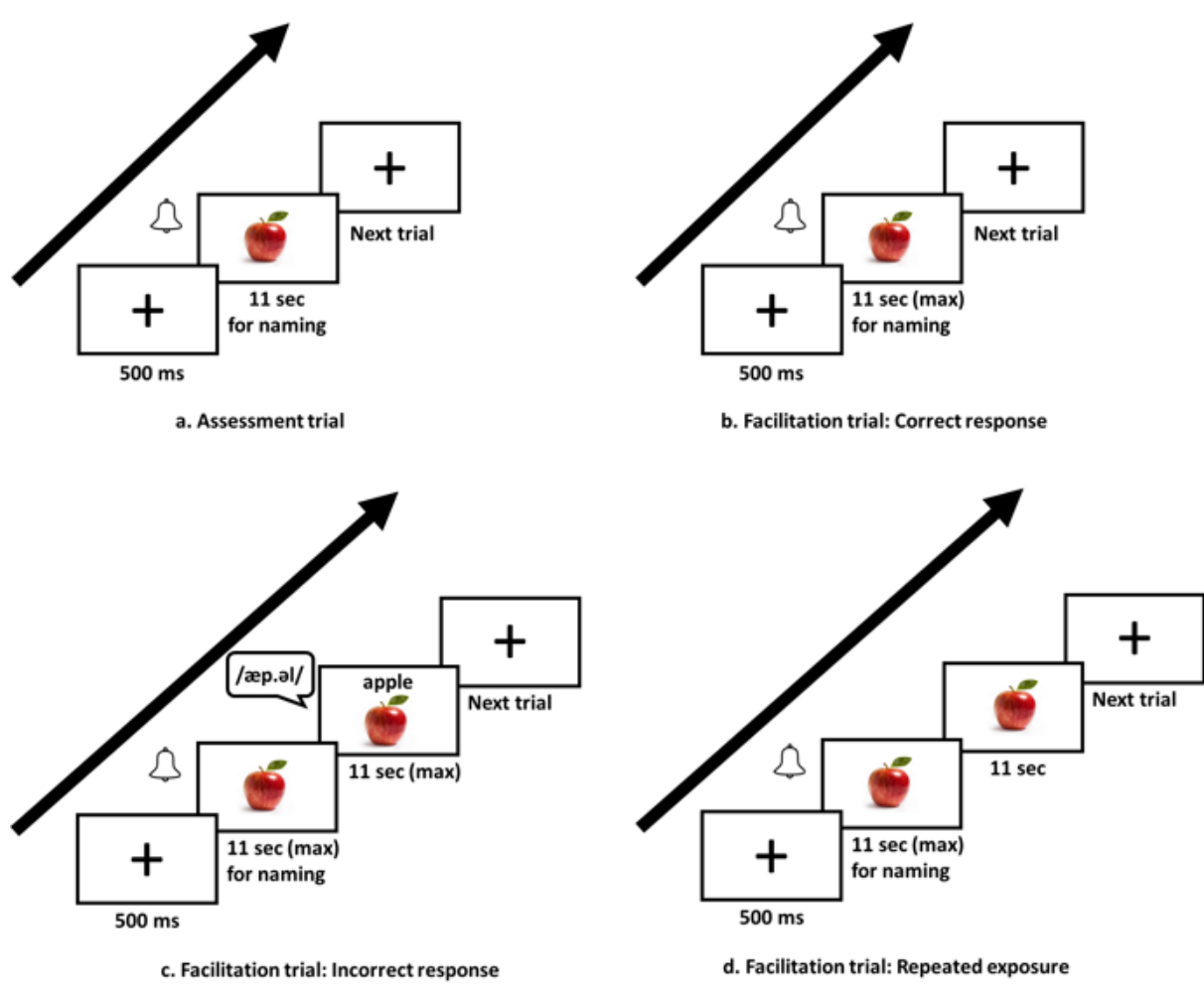
3.2.4.4 Picture naming trials during facilitation

The sequence of each picture naming trial is graphically presented in Figure 3.2. Every trial commenced with a fixation cross, directing attention towards the centre of the screen. Next, an alerting sound and a centrally-positioned picture were simultaneously presented. The participant would have a maximum of 11 seconds to name the picture. If a correct spoken response was produced within 11 seconds, the experimenter would press a button, which moved the participant directly to the next trial (Figure 3.2b). If an incorrect or no response was provided, the picture stayed online for the full 11-seconds, before it was presented again with the facilitation cue assigned (Figure 3.2c), i.e., with the orthographic cue printed at the top (either the first letter or the entire word, for conditions 1 and 2) or with both orthographic and phonological cues simultaneously provided (either both initial letter and phoneme or both written and spoken word in its entirety, for conditions 3 and 4). The picture alongside its orthographic facilitation cue would remain on screen for an additional 11 seconds. The phonological cue was only played once throughout this 11-second duration. For repeated exposure, the sequence was similar to that of the facilitation trials, except that no facilitation cues were presented (Figure 3.2d). The same picture was displayed again for an extra 11 seconds. Across the facilitation trials (Figures 3.2b to d), there is thus no explicit feedback on the accuracy of each response. However, because the trial moves forward to the next item once a correct response is given, the feedback is thus implicit.

Each condition (the four facilitation cues and repeated exposure; conditions 1 to 5) was presented as a block to reduce confusion. In order to counter transfer effects between the conditions, participants went through the five conditions in a sequence counter-balanced using Latin-square. Finally, five practice trials were also provided at the start of each block. Only items participants successfully named twice on both baseline sessions and not selected as 'items for encouragement' during facilitation, were used as practice items. The practice trials served two functions: To (1) familiarise the participant with the facilitation procedure, and (2) act as additional audio checks. Based on the audio clips played during the practice trials, the sound volume was further adjusted to the comfort level indicated by each participant.

All sounds were played over a portable USB speaker (HONK HK-5002) for clarity. Throughout the study, naming trials were run using PsychoPy (Peirce et al., 2019) on a Toshiba Portege laptop.

Figure 3.2. Format of assessment and facilitation trials.



3.2.4.5 Data extraction and fidelity check

All spoken responses during assessments were digitally audio-recorded using the Sony ICDPX370 Mono Digital Voice Recorder.

Data were extracted from the audio recordings at the item-level for each participant and for each session. Three main outcome measures of picture naming were first coded in their raw forms. They were operationalised as follow.

- (1) Initial naming accuracy: This referred to whether the first complete response during the 11-second naming interval was correct, excluding fillers and hesitations (e.g., 'hmm', 'errr').
- (2) Final naming accuracy: This referred to whether the last complete attempt during the 11-second naming interval was correct, excluding fillers and hesitations. Both initial and final naming accuracies were therefore binary variables. Speech sound distortions were not counted as errors. Responses were only scored as 'correct' if they were either identical to the target word based on the picture norms, or were acceptable alternatives (e.g., 'brolly' for 'umbrella').
- (3) RT: This referred to the time taken to produce the correct response. RT was extracted using Audacity 2.3.0 (Audacity Team, 2018), and it was measured from the onset of the trial to the onset of the first phoneme of the target word.

An external rater fluent in English rated 20% of the data independently; this 20% was sampled by way of randomly selecting one participant's full dataset using Research Randomiser. The order of the sessions was additionally randomised again using Research Randomiser, before they were sent for external rating. The external rater was therefore blind to the conditions and the order of the sessions of each item. For every item, an accuracy score was considered as agreed upon if the pair of independent coding from the external rater and the original scorer was identical. For RT, agreement was determined in two steps: (1) For each rater and within each session, the RT for every item should fall within 3 *SDs* of the average timings¹⁹ computed across items for that rater. This is to check for intra-rater outliers. (2) Across the two raters and within the same session, the averaged RT for each item should also fall within 3 *SDs* of the collective averaged timings computed across items *and* across raters. This is to account for interrater discrepancies. Disagreements on any items were resolved through discussions. Overall, interrater agreement was high across all measures, with at least 90% percentage agreement on each measure for each assessment session [Average percentage agreement across sessions: Initial naming accuracies = 99.72%; Final naming accuracies = 99.58%; RT = 100%] and with Krippendorff's α on each measure per session exceeding .800 [Average α -value across sessions: Initial naming accuracies = .989; Final naming accuracies = .988; RT = 1.00](Krippendorff, 2004; Freelon, 2010, 2013). Finally, for RT, the average time across the two raters was computed for each item; these were then used in subsequent RT analyses.

¹⁹ The number of *SDs* is arbitrary, but 3 *SDs* is in line with the procedural standard applied in psycholinguistic studies (usually 2.5 *SDs* to 3 *SDs*, e.g., Balota and colleagues, 2007, applied 3 *SDs*, while Sze, Rickard Liow, and Yap, 2014, was more stringent, opting 2.5 *SDs* as the cut-off). At the higher end, this bandwidth should be able to take into account the likelihood of heterogeneity when working with individuals with language disorders.

3.2.4.6 Group-level analyses

3.2.4.6.1 Group-level analyses for primary research questions (1) to (3): Outcomes of cue conditions across and between time points

For primary research questions (1) to (3), mean proportional outcome was computed for each of the three outcome measures (initial and final naming accuracies, RT), and these were entered as dependent variables. The values were computed at the participant level, and each value was obtained by averaging the respective accuracy scores and RT across items for each condition at each assessment time point. Mean proportion was used instead of raw accuracy values, because different participants received different number of treated items during facilitation. Proportion scores were similarly used in group analyses by other researchers in the same context (e.g., pp. 155-156 in Best et al., 2002).

Analyses then proceeded using non-parametric approaches, given that there were only five participants (Siegel & Castellan, Jr., 1988). Six independent sets of analyses were run, corresponding to the six facilitation/control conditions. Within each of these six sets, analyses were also separately conducted for the three outcome measures.

Within each set of analysis, Friedman Analysis of Variance (ANOVA) was first used to compare performances collectively across the four time points, on each outcome measure. As follow up, three post hoc comparisons using the Wilcoxon signed ranks test were also planned. The main purpose of these planned comparisons was to determine whether the significant findings from Friedman ANOVA simply reflected a general trend of improvement across the four time points, or whether a significant change arose, owing to a definite boost from the cue administered.

The first pairwise comparison examined whether the two baselines were significantly different. This was to rule out large discrepancies between baselines, which might otherwise explain the significant Friedman ANOVA result. For the second and third pairwise comparisons, performances at the 'Post-Facilitation Assessments 1' and '2' were compared respectively against the naming performances at 'Baseline 2'. Baseline 2 represented the most recent baseline performance prior to cueing. These last two pairwise comparisons would clarify the time course of the facilitation effects.

Finally, LSD procedure was applied to the post hoc tests, firstly in view of the small sample ($n = 5$ participants) and secondly, it would control for Type I errors associated with making three or even four pairwise comparisons, while avoiding Type II errors (Green & Salkind, 2005, pp. 417-418; for a more thorough explanation, see Howell, 2014, pp. 416-420).

3.2.4.6.2 Group-level correlational analyses for secondary research question (4): General language processes

To investigate secondary research question (4), the variable ‘facilitation proportional gain’ (referred to as ‘facilitation gain’ for brevity) was first calculated (see p. 294 in Lambon Ralph et al., 2010, for a similar operationalisation). Computation for ‘facilitation gain’ was processed at the participant level separately for the three naming outcome measures at each post-facilitation time point and for each cue condition. For accuracy measures, the formula for ‘facilitation gain’ was: Subtract naming score at baseline session 2 from the total naming score obtained post-therapy, and divide that by the total number of facilitation items still not successfully named by the final baseline. For instance, if there were 30 items assigned for facilitation, the participant named 6 correctly at the baseline session 2, and 10 correctly a week after completing the facilitation session, calculation for ‘facilitation gain’ one-week post-intervention would be: $(10-6)/(30-6) = 0.1667$. For RT, if the participant’s average RT across items at the baseline session 2 was 9800 ms, and average RT one-week post-therapy was 4200 ms, then the ‘facilitation gain’ at one week would be: $(9800 - 4200)/9800 = 0.5714$. To summarise, these numbers indicated the amount of improvement the participant achieved following the facilitation cue, and that this improvement was proportional to the maximum change possible given the facilitation set size/baseline RT.

Next, again at the participant level, three overall baseline language scores were computed to index the three general language processing abilities of interest. The three general language processing abilities would correspond to the three processing areas that anchor most psycholinguistic models, namely the processing of: (1) written language (orthographic input), (2) spoken language (phonological input) and (3) semantics (e.g., Seidenberg & McClelland, 1989; Kay, Lesser, & Colheart, 1996; also see section 1.3). In this thesis, written language processing refers to the processes stemming from written input that result in spoken word production; spoken language processing will be the analogous processes but with auditory modality as the input. Semantic processing will be the processes involved in decoding meaning, be it at the conceptual and/or lexical level(s). Written and spoken language processes also broadly align with the skills required to engage with orthographic and phonological cues to support successful spoken naming.

Overall scores were used to simplify analyses as they functioned as a summary index for a specific language processing ability. Functionally, having a combined score also meant data from all available baseline tasks could be utilised, even when an assessment contained few test items (e.g., 10 items in ‘CAT 2: Semantic Memory’). An overall score (an

average z-score) was thus computed per participant for each general language process following these steps:

Step (a): For each assessment task, compute its mean and standard deviation by pooling together the five participants' baseline performances on that test.

Step (b): For each participant, a standardised z-score was computed per assessment using the mean and standard deviation calculated for that assessment in step (a).

Step (c): For each participant, an average z-score was then computed across the assessments relevant to that particular language processing ability. Take for instance P1's 'written language processing ability', an average z-score was computed across P1's standardised scores from the written language assessments he completed at baseline. The breakdown of assessments within each processing thread is stated below and presented in Table 3.7. The same steps were repeated to obtain overall scores for P1's 'spoken language processing' and 'semantic processing' abilities, i.e., compute average z-scores across the spoken language and semantic processing tasks respectively.

The assessments used to compute the 'written language processing' abilities were: Letter sounding (PALPA 22a), letter naming (PALPA 22b), visual lexical decision task (PALPA 25), reading words (CAT 20), reading complex words (CAT 21), reading function words (CAT 22), and reading nonwords (CAT 23). The tasks for 'spoken language processing' abilities were the two nonword repetition tasks (CAT 14 and PALPA 8) and the auditory lexicon decision task (PALPA 5); and for 'semantic processing', the tasks were semantic memory (CAT 2), comprehension of spoken words (CAT 7), comprehension of written words (CAT 8), auditory synonym judgement (PALPA 49) and written synonym judgement (PALPA 60).

Finally, to explore which of these general language processes (written language, spoken language and semantic) were tapped on to extract cue benefits, the participants' averaged standardised scores for each language processing ability were correlated with their 'facilitation gains' on initial/final naming accuracies and RT, for each cue condition. Kendall's rank-order correlation was used due to the small sample size (see Siegel & Castellan Jr., 1988). No rank-order correlations were computed between the averaged standardised scores for spoken language processing and the two 'orthography only' cue conditions because there was no theoretical motivation to associate auditory input processing with visual word form decoding.

3.2.4.6.3 Group-level correlational analyses for secondary research question (4): Specific language skills

To examine the specific language skills recruited to successfully exploit a specific cue condition, ‘facilitation gains’ were correlated with the proportional accuracy scores for specific language assessments administered at baseline. Again, Kendall rank-order correlation was applied. Proportional accuracy was computed by dividing the participant’s score obtained at baseline by the maximum score for that assessment. For example, the maximum score for ‘PALPA 22: Letter Sounding’ is ‘26’, and had the participant scored ‘20’ at baseline, his proportional accuracy for ‘Letter Sounding’ would be: $20/26 = 0.7692$. Proportional accuracy scores were used instead of raw accuracy values, because the raw values were not directly comparable since maximum scores differed across assessments.

For cue conditions with an orthographic component, the following specific language processing skills and their tasks (stated within parentheses) were of interest. Baseline performances on these tasks were therefore correlated with the three indices of responsiveness towards each cue condition.

- (1) Written lexical-semantic processing (PALPA 50: Written Synonym Judgement; 60 items),
- (2) Written lexical processing (PALPA 25: Visual Lexical Decision Task, 120 items), and
- (3) Written sub-lexical processing (PALPA 22: Letter Sounding, 26 items; PALPA 36: Nonword Reading: 24 items)
- (4) Written word processing (CAT 20: Reading of Simple Words, 24 items)²⁰: This involves the above three specific written language processing skills.

As for conditions incorporating phonological cues, the interest was in whether the following specific skills would also be harnessed to profit from the multimodal cues. Baseline assessment scores from these tasks were correspondingly added to the correlation matrix.

- (5) Spoken lexical-semantic processing (PALPA 49: Auditory Synonym Judgement, 60 items),
- (6) Spoken lexical processing (PALPA 5: Auditory Lexical Decision Task, 160 items), and
- (7) Sound-to-phoneme conversion (PALPA 8: Nonword Repetition, 30 items)

In Table 3.7, the tasks are also classified according to the specific language skill. Since this set of correlations involved scores from solitary assessments (i.e., not combined scores across assessments as per ‘general language processes’ in Section 3.2.4.6.1), care was taken to ensure that the assessments included contained a sufficient spread of items for adequate representation (smallest $n = 24$ items) and to allow abilities to be properly distinguished. This takes on board the need for discriminating tests for clearer correlation

²⁰ From the entire list of baseline assessments, ‘CAT 21: Reading complex words’ was dropped, because all participants scored zero on the task. ‘CAT 23: Reading nonwords’ was omitted in favour of its duplicate, ‘PALPA 36: Nonword reading’ because the latter comprised of more test stimuli.

outcomes (discussion of a similar issue can be found on p. 2349 and p. 2352 in Best et al., 2013)²¹. Similar to the above set of rank-order correlations (Section 3.2.4.6.1) and to mitigate spurious outcomes, proportional accuracies from auditory tasks were not correlated with the facilitation gains for ‘orthography only’ cues, as there was no theoretical basis to draw sensible mutual links.

3.3 Results

Results are reported in the order of the research questions. Only results from the group-level analyses are provided here. Outcomes from individual-level analyses will be covered in Chapter 4.

3.3.1 Primary research questions (1) to (3): At the participant group level, which unit and mode of cueing will result in significant improvement; also, how durable are the facilitation effects of each combination of cue (unit x modality)?

Figure 3.3 summarises the descriptive statistics for all three outcome measures. The full set of results using Friedman analyses and post hoc Wilcoxon tests are presented in Tables 3.2 to 3.4. To assist interpretation, cells containing results that are significant and those approaching significance are highlighted in yellow.

As a preliminary check, it should be pointed out that performances between the two baselines during post hoc tests (conducted where applicable) were not significantly different on all three outcome measures (Tables 3.3 - 3.5). These suggested that the overall performance across participants prior to facilitation was relatively stable.

Both ‘Initial Letter: Orthography’ and ‘Whole Word: Orthography’ cues produced significant outcomes on all three naming measures based on Friedman ANOVA. 10 out of the possible 12 post hoc results that compared post-facilitation outcomes against the respective final baselines were significant (Tables 3.3 to 3.5). On average, participants’ naming improved following both types of ‘orthography only’ cues at the two post-facilitation

²¹ The comprehension tasks (CAT 2: Semantic Memory, 10 items; CAT 7: a spoken word-to-picture task, 15 items; CAT 8: a written word-to-picture task, 15 items) were therefore excluded to avoid specious outcomes. In any case, subsequent analyses showed that semantic processes at the general language level failed to correlate with the cue outcomes. The repetition tasks and some reading tests in CAT were also thus excluded for their small sets of test items (e.g., CAT 22: Reading Function Words, 3 items). ‘PALPA 22: Letter Naming’ was also not included although there were 26 items, because theoretically, it is not the ability to name letters that would be pertinent to the use of written cues but letter sounding.

time points (15 minutes and one week later) on all three measures, apart from final naming accuracy with the 'Initial Letter: Orthography' cue, where effects failed to reach significance 15 minutes immediately post-facilitation [$p = .11$; Table 3.4].

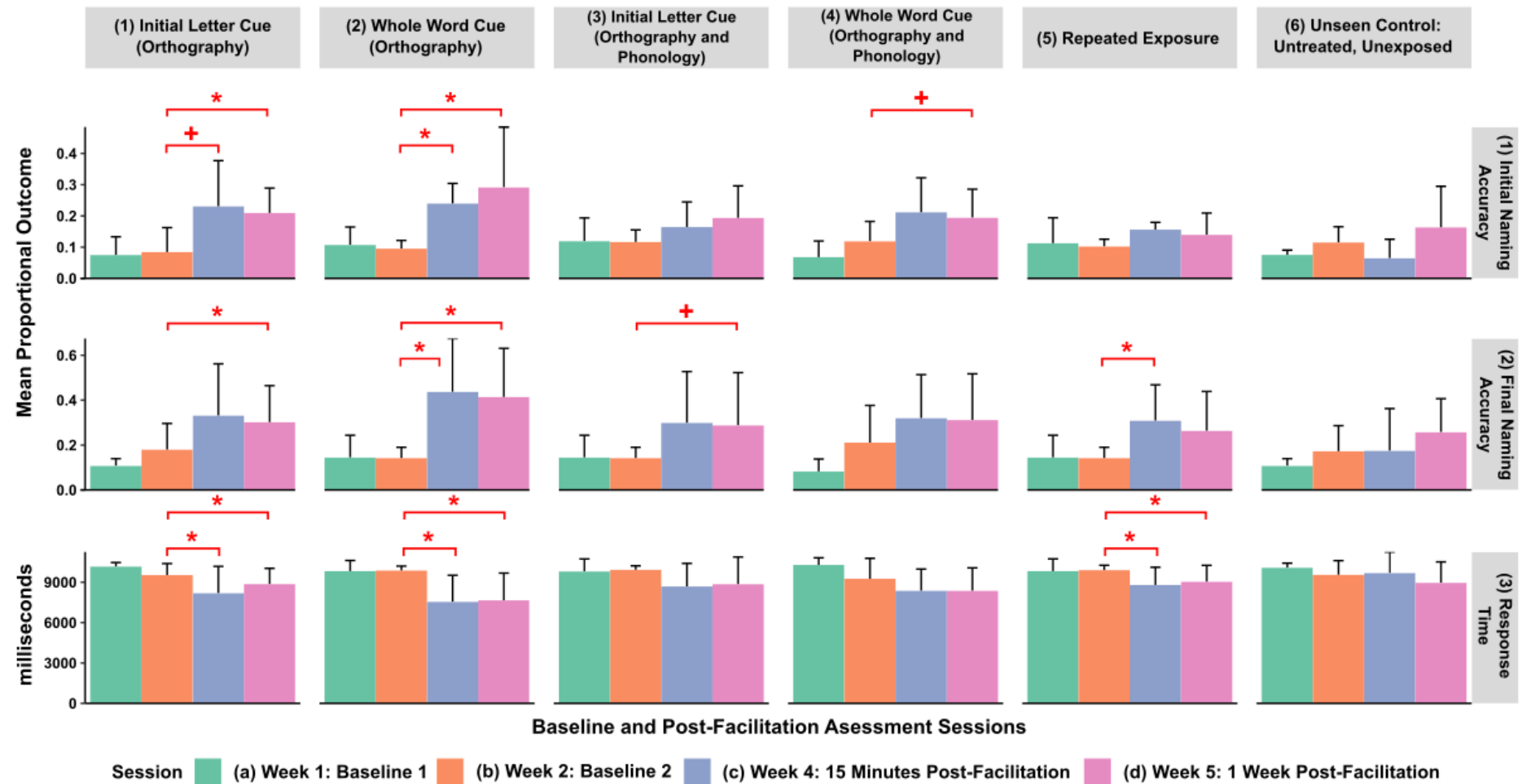
Turning to the results of multimodal cues, facilitation effects were marginally significant on only two occasions, i.e., (1) Effects from 'Whole Word: Orthography and Phonology' cue approached significance on initial naming accuracy when assessed one week later [$p = .059$; Table 3.3], and (2) 'Initial Letter: Orthography and Phonology' marginally improved final naming accuracy one week later [$p = .066$; Table 3.4]. The rest of the results were not significant. Interestingly, neither multimodal cue enhanced the participants' average naming speed (Table 3.5).

Effects were also observed for the 'Repeated Exposure' condition on two measures, i.e., final naming accuracy and response times. To verify that the significant findings reported earlier for cue conditions on these two outcome measures were not simply due to participants' repeated experience seeing/attempting to name the pictures, additional tests were run. Mean proportional outcomes from cue conditions whose results were significant on these two measures, were now compared against data from 'Repeated Exposure' separately at the two post-facilitation time points, using Wilcoxon signed ranks tests.

Results from these additional Wilcoxon tests revealed that, on all three outcome measures, only facilitation effects following the 'Whole Word: Orthography' cue, were significantly stronger than the corresponding naming performances induced by 'Repeated Exposure' [Initial naming accuracy (15 minutes later): $z = -1.83$, $p = .068$, $r = -0.58$; Initial naming accuracy (one week later): $z = -2.02$, $p = .043$, $r = -0.64$; Final naming accuracy (15 minutes later): $z = -2.04$, $p = .041$, $r = -0.65$; Final naming accuracy (one week later): $z = -2.03$, $p = .042$, $r = -0.64$; RT (15 minutes and one week later): $z = -2.02$, $p = .043$, $r = -0.64$ in both instances].

In contrast, performances for the remaining cue conditions did not significantly differ from what was achieved based on 'Repeated Exposure' alone at the two post-facilitation time points [all ps at least $> .10$]. To check that the baseline performances at Week 2 were comparable between cue conditions (that yielded significant effects) and the repeated exposure condition, corresponding Wilcoxon tests were conducted on final naming accuracy and RT collected at Week 2; all comparisons were not significant [all $ps > .079$]. None of the omnibus Friedman ANOVA results was significant for the unseen control condition either [$ps > .15$; Tables 3.3 to 3.5]. All the above results are summarised in Table 3.6.

Figure 3.3. Descriptive statistics of group-level results ($n = 5$) for the six conditions across four sessions on each of the three naming outcome measures.



+ .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

Table 3.3. Results of the group analysis by-participants based on the mean proportional accuracy of their initial naming responses.

Condition	Friedman ANOVA	Baseline 2 vs. Baseline 1	15 Minutes Post- Facilitation vs. Baseline 2	1-Week Post-Facilitation vs. Baseline 2
Initial Letter: Orthography	* $\chi^2(3, N = 5) = 9.00$, $p = .029$, Kendall's $W = .60$	<i>n.s.</i> $z = -0.38$, $p = .71$, $r = -0.12$	† $z = -1.83$, $p = .068$, $r = -0.58$	* $z = -2.06$, $p = .039$, $r = -0.65$
Whole Word: Orthography	** $\chi^2(3, N = 5) = 12.83$, $p = .005$, Kendall's $W = .86$	<i>n.s.</i> $z = 0$, $p = 1.00$, $r = 0$	* $z = -2.03$, $p = .042$, $r = -0.64$	* $z = -2.03$, $p = .042$, $r = -0.64$
Initial Letter: Orthography + Phonology	<i>n.s.</i> $\chi^2(3, N = 5) = 4.06$, $p = .25$, Kendall's $W = .27$			
Whole Word: Orthography + Phonology	† $\chi^2(3, N = 5) = 7.36$, $p = .061$, Kendall's $W = .49$	<i>n.s.</i> $z = -1.63$, $p = .10$, $r = -0.52$	<i>n.s.</i> $z = -1.63$, $p = .10$, $r = -0.52$	† $z = -1.89$, $p = .059$, $r = -0.60$
Repeated Exposure	<i>n.s.</i> $\chi^2(3, N = 5) = 5.65$, $p = .13$, Kendall's $W = .38$			
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $\chi^2(3, N = 5) = 5.31$, $p = .15$, Kendall's $W = .35$			

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

Table 3.4. Results of the group analysis by-participants based on the mean proportional accuracy of their final naming responses.

Condition	Friedman ANOVA	Baseline 2 vs. Baseline 1	15 Minutes Post- Facilitation vs. Baseline 2	1-Week Post-Facilitation vs. Baseline 2
Initial Letter: Orthography	* $\chi^2(3, N = 5) = 9.52$, $p = .023$, Kendall's $W = .63$	<i>n.s.</i> $z = -1.51, p = .13$, $r = -0.48$	<i>n.s.</i> $z = -1.60, p = .11$, $r = -0.51$	* $z = -2.04, p = .041$, $r = -0.65$
Whole Word: Orthography	** $\chi^2(3, N = 5) = 12.12$, $p = .007$, Kendall's $W = .81$	<i>n.s.</i> $z = -2.76, p = .78$, $r = -0.087$	* $z = -2.03, p = .042$, $r = -0.64$	* $z = -2.03, p = .042$, $r = -0.64$
Initial Letter: Orthography + Phonology	† $\chi^2(3, N = 5) = 7.53$, $p = .057$, Kendall's $W = .50$	<i>n.s.</i> $z = -2.76, p = .78$, $r = -0.087$	<i>n.s.</i> $z = -1.47, p = .14$, $r = -0.47$	† $z = -1.84, p = .066$, $r = -0.58$
Whole Word: Orthography + Phonology	<i>n.s.</i> $\chi^2(3, N = 5) = 6.06$, $p = .11$, Kendall's $W = .40$			
Repeated Exposure	* $\chi^2(3, N = 5) = 10.66$, $p = .014$, Kendall's $W = .71$	<i>n.s.</i> $z = -2.76, p = .78$, $r = -0.087$	* $z = -2.06, p = .039$, $r = -0.65$	<i>n.s.</i> $z = -1.60, p = .11$, $r = -0.51$
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $\chi^2(3, N = 5) = 5.30$, $p = .15$, Kendall's $W = .35$			

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

Table 3.5. Results of the group analysis by-participants based on the average naming response times.

Condition	Friedman ANOVA	Baseline 2 vs. Baseline 1	15 Minutes Post- Facilitation vs. Baseline 2	1-Week Post-Facilitation vs. Baseline 2
Initial Letter: Orthography	* $\chi^2(3, N = 5) = 9.96,$ $p = .019$, Kendall's $W = .66$	<i>n.s.</i> $z = -1.48, p = .14,$ $r = -0.47$	* $z = -2.02, p = .043,$ $r = -0.64$	* $z = -2.02, p = .043,$ $r = -0.64$
Whole Word: Orthography	** $\chi^2(3, N = 5) = 12.60,$ $p = .006$, Kendall's $W = .84$	<i>n.s.</i> $z = -0.13, p = .89,$ $r = -0.043$	* $z = -2.02, p = .043,$ $r = -0.64$	* $z = -2.02, p = .043,$ $r = -0.64$
Initial Letter: Orthography + Phonology	<i>n.s.</i> $\chi^2(3, N = 5) = 4.44,$ $p = .22$, Kendall's $W = .30$			
Whole Word: Orthography + Phonology	<i>n.s.</i> $\chi^2(3, N = 5) = 6.36,$ $p = .095$, Kendall's $W = .42$			
Repeated Exposure	** $\chi^2(3, N = 5) = 12.12,$ $p = .007$, Kendall's $W = .81$	<i>n.s.</i> $z = -0.13, p = .89,$ $r = -0.043$	* $z = -2.02, p = .043,$ $r = -0.64$	* $z = -2.02, p = .043,$ $r = -0.64$
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $\chi^2(3, N = 5) = 3.96,$ $p = .27$, Kendall's $W = .26$			

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

Table 3.6. Summary of findings for the cue and repeated exposure conditions for group analyses for each outcome measure.

Condition		<u>Initial Letter Cue: Orthography</u>			<u>Whole Word Cue: Orthography</u>			<u>Initial Letter Cue: Orthography and Phonology</u>			<u>Whole Word Cue: Orthography and Phonology</u>			<u>Repeated Exposure</u>		
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT
Group (Analysis by Participants)		*	*	*	**	**	**	<i>n.s.</i>	†	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	**
Post-hoc (against Baseline 2)	15 mins Post- vs. BL2	†	<i>n.s.</i>	*	*	*	*	–	<i>n.s.</i>	–	<i>n.s.</i>	–	–	–	*	*
	x1 Week Post- vs. BL2	*	*	*	*	*	*		†		†				<i>n.s.</i>	*
Post-hoc (against rep. exposure)	Final Baselines	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>		<i>n.s.</i>		<i>n.s.</i>					
	15 mins Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	†	*	*	–	<i>n.s.</i>	–	<i>n.s.</i>	–	–			
	x1 Week Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	*	*		<i>n.s.</i>		<i>n.s.</i>					

IN Acc.: Initial Naming Accuracy, FN Acc.: Final Naming Accuracy, Pts: Participants

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

– : Post hoc not required as effects of ‘Repeated Exposure’ were not significant, based on its overall group analyses

Lastly, for readers interested in understanding the scale of the changes in detail to facilitate their session planning, the full set of mean numerical changes across participants in each condition, is published in Appendix 3.4. To illustrate and given how robust results were for the ‘Whole Word: Orthography’ cue (Table 3.6), raw mean values of this prompt’s improvements relative to baseline performances at Week 2, are highlighted as follow: When assessed 15 minutes post-facilitation, on average the participants improved by 3.60 items ($SD = 1.95$; reflecting a proportional change of 0.16) for initial naming accuracy and 6.80 items ($SD = 2.95$; proportional change = 0.35) for final naming accuracy, compared to the respective final baseline outcomes at Week 2. Participants’ responses were also on average, faster by 2327.11 ms ($SD = 1766.28$; proportional change = 0.24). When assessed a week later, overall participants still improved by 4.20 items ($SD = 3.03$; proportional change = 0.21) for initial naming accuracy and 6.40 items ($SD = 2.61$; proportional change = 0.33) for final naming accuracy. In addition, their naming remained faster by 2210.74 ms ($SD = 1769.87$; proportional change = 0.23).

3.3.2 Secondary research question (4): At the participant group level, what is the relationship between baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?

The full set of Kendall’s rank-order correlations between the outcomes from the four cue conditions at the two time points and baseline language performances (overall average z-scores across assessments and proportional accuracy scores from specific assessments) are published in Table 3.7. The findings here are exploratory. Significant results are highlighted in yellow to ease reading.

In terms of general language processes, when assessed 15 minutes immediately post-facilitation, ‘written language processing’ abilities were significantly and strongly correlated with the final naming outcome from ‘Whole Word: Orthography Only’ cues, such that better overall ‘written language processing’ abilities were associated with more accurate final naming responses [final naming accuracy: $r_b > .96$, $p < .001$].

Similar strong and significant positive relationships between overall ‘written language processing’ abilities and other cue conditions were also observed at one-week post-facilitation – these included ‘written language processing’ abilities and the outcomes from ‘Initial Letter: Orthography Only’ cues [final naming accuracy: $r_b = .95$, $p = .023$; RT: $r_b > .96$, $p < .001$], as well as with the initial naming outcome from ‘Whole Word: Orthography and Phonology’ cues [initial naming accuracy: $r_b > .96$, $p < .001$]. Marginally significant correlations were also identified elsewhere (see Table 3.7), notably between ‘written

Table 3.7. Kendall rank-order correlations between scores based on general language processes or specific baseline language assessments and the proportional naming gains yielded at different post-facilitation time points (i.e., Initial Naming (IN): Therapy Gain 15 minutes and 1 week later (15 Mins, 1 Wk), Final Naming (FN): Therapy Gain 15 minutes and 1 week later, and Response Time (RT): Therapy Gain 15 minutes and 1 week later) for each cue condition.

Initial Letter Cue: Orthography Only							Whole Word Cue: Orthography Only					
	IN: 15 Mins	FN: 15 Mins	RT: 15 Mins	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk	IN: 15 Mins	FN: 15 Mins	RT: 15 Mins	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk
General Language (Lang.) Processes - Average z-Scores across tasks												
Written Lang. Processing	.20	.74 (<i>p</i> = .077)	.80 [†] (<i>p</i> = .05004)	.40	.95* (<i>p</i> = .023)	1.00***	.20	1.00***	.80 [†] (<i>p</i> = .05004)	.60	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)
Spoken Lang. Processing	-	-	-	-	-	-	-	-	-	-	-	-
Semantic Lang. Processing	-.20	.11	0	.40	.32	.20	-.20	.20	0	.20	.40	.40
Written Lexical-Semantic Processing												
Written Synonym Judgement (PALPA 50)	-.20	.32	.40	0	.53	.60	.20	.60	.40	.60	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)
(Written) Lexical Processing												
Visual Lexical Decision (PALPA 25)	0	.32	.20	.20	.53	.40	-.40	.40	.20	0	.60	.60
(Written) Sub-Lexical Processing												
Letter Sounding (PALPA 22)	0	-.32	-.20	.20	-.11	0	0	0	.20	.40	.20	.20
Nonword Reading (PALPA 36)	.32	.33	.32	.63	.67	.63	0	.63	.63	.63	.63	.63
Written Word Processing												
Word Reading (CAT 20: Simple Words)	.20	.74 (<i>p</i> = .077)	.80 [†] (<i>p</i> = .05004)	.40	.95* (<i>p</i> = .023)	1.00***	.20	1.00***	.80 [†] (<i>p</i> = .05004)	.60	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)
Spoken Lexical-Semantic Processing												
Auditory Synonym Judgement (PALPA 49)	-	-	-	-	-	-	-	-	-	-	-	-
(Spoken) Lexical Processing												
Auditory Lexical Decision (PALPA 5)	-	-	-	-	-	-	-	-	-	-	-	-
Sound-to-Phoneme Conversion												
Nonword Repetition (PALPA 8)	-	-	-	-	-	-	-	-	-	-	-	-

† .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001

(continued)

Initial Letter Cue: Orthography and Phonology							Whole Word Cue: Orthography and Phonology					
	IN: 15 Mins	FN: 15 Mins	RT: 15 Mins	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk	IN: 15 Mins	FN: 15 Mins	RT: 15 Mins	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk
General Language (Lang.) Processes - Average z-Scores across tasks												
Written Lang. Processing	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)	.40	.40	.20	.53	0	0	1.00***	.20	.20
Spoken Lang. Processing	-.40	-.80 [†] (<i>p</i> = .05004)	-.80 [†] (<i>p</i> = .05004)	-.40	0	.20	-.11	.40	.40	-.60	.20	.20
Semantic Lang. Processing	.40	.40	.40	0	.80 [†] (<i>p</i> = .05004)	.60	-.32	-.40	-.40	.20	.20	.20
Written Lexical-Semantic Processing												
Written Synonym Judgement (PALPA 50)	.80 [†] (<i>p</i> = .05004)	.40	.40	.40	.80 [†] (<i>p</i> = .05004)	.60	.11	0	0	.60	.60	.60
(Written) Lexical Processing												
Visual Lexical Decision (PALPA 25)	.60	.60	.60	-.20	.20	0	.32	.20	.20	.40	.40	.40
(Written) Sub-Lexical Decoding												
Letter Sounding (PALPA 22)	.20	-.20	-.20	.60	.60	.80 [†] (<i>p</i> = .05004)	-.53	-.60	-.60	0	0	0
Nonword Reading (PALPA 36)	.63	.63	.63	.63	.63	.63	-.17	-.63	-.63	.63	0	0
Written Word Processing												
Word Reading (CAT 20: Simple Words)	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)	.40	.40	.20	.53	0	0	1.000**	.20	.20
Spoken Lexical-Semantic Processing												
Auditory Synonym Judgement (PALPA 49)	.40	.40	.40	0	.80 [†] (<i>p</i> = .05004)	.60	-.32	-.40	-.40	.20	.20	.20
(Spoken) Lexical Processing												
Auditory Lexical Decision (PALPA 5)	0	-.40	-.40	-.40	0	-.20	.32	.80 [†] (<i>p</i> = .05004)	.80 [†] (<i>p</i> = .05004)	-.20	.60	.60
Sound-to-Phoneme Conversion												
Nonword Repetition (PALPA 8)	-.89 [†] (<i>p</i> = .037)	-.89 [†] (<i>p</i> = .037)	-.89 [†] (<i>p</i> = .037)	-.22	-.45	-.22	-.35	0	0	-.89 [†] (<i>p</i> = .037)	-.22	-.22

† .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001

language processing' and 'Initial Letter: Orthography and Phonology' on all three outcome measures at 15 minutes post-facilitation [in all three instances: $r_b = .80$, $p = .05004$].

Correlations with the remaining two general language processes were not significant, although some approached significance: i.e., (1) the relationships between 'spoken language processing' and 'Initial Letter: Orthography and Phonology' [final naming accuracy and RT at 15 minutes post-facilitation: $r_b = -.80$, $p = .05004$ in both], and 'semantic processing' with 'Initial Letter: Orthography and Phonology' at one-week post-facilitation [final naming accuracy: $r_b = .80$, $p = .05004$].

Turning to specific language skills, participants with higher scores on 'Word Reading (CAT 20)' were associated with better outcomes on at least three facilitation cues. These included the strong and significant positive correlations between performance on word reading and outcomes from (1) 'Whole Word: Orthography Only' when assessed 15 minutes post-facilitation [final naming accuracy: $r_b > .96$, $p < .001$], (2) 'Initial Letter: Orthography Only' cues when assessed one-week later [final naming accuracy: $r_b = .95$, $p = .023$; RT: $r_b > .96$, $p < .001$] and (3) 'Whole Word: Orthography and Phonology' cues when assessed one-week later [initial naming accuracy: $r_b > .96$, $p < .001$].

There were significant inverse relationships between 'Nonword Repetition (PALPA 8)' and outcomes based on two multimodal cues, such that participants with better nonword repetition performance profited less from these cue types. For the 'Initial Letter: Orthography and Phonology' cue, significant negative associations were found between all three of its outcome measures and nonword repetition when assessed 15-minutes later: $r_b = -.89$, $p = .037$ each; for the 'Whole Word: Orthography and Phonology' cue, a significant negative correlation was identified between the outcome based on initial naming accuracy and the performance on nonword repetition, when re-assessed one-week later [$r_b = -.89$, $p = .037$].

Several other correlations also approached significance (Table 3.7); notably between the baseline performance on 'Word Reading (CAT 20)' and outcomes from 'Initial Letter: Orthography and Phonology' cue, which was replicated across all three outcome measures at 15-minutes post-facilitation [$r_b = .80$, $p = .05004$]. Incidentally, outcomes from 'Initial Letter: Orthography and Phonology' cue were also marginally correlated with the proportional scores on three other tasks in the positive direction, and these were: (1) written synonym judgement with initial naming accuracy 15-minutes post-facilitation and final naming accuracy at one week afterwards, (2) auditory synonym judgement with final naming

accuracy one week afterwards, and (3) letter sounding with RT one-week post-facilitation [$\tau_b = .80$, $p = .05004$ in each of these cases].

3.6.3 Control task

Group analysis of accuracy scores from the nonword repetition tasks using Wilcoxon signed ranks tests showed no significant change in nonword repetition performance at one-week post-facilitation [Mean outcome across participants: At baseline = 11.40 ($SD = 5.13$), one-week post-facilitation = 12.00 ($SD = 5.61$); baseline vs. one-week post-facilitation: $z = -0.74$, $p = .46$, $r = -0.33$]. These results further corroborate that effects reported earlier, were likely specific to the facilitation cues.

3.4 Discussion

This facilitation study investigated the features (unit, modality) of an optimal orthographic cue (Primary research questions 1 and 2) and the duration of their attendant effects (Primary research question 3). Two variables ('Unit of cue: Initial Letter vs. Whole Word' x 'Modality: Orthography vs. Orthography and Phonology') were crossed to create four cue conditions, and their facilitation outcomes were tracked across two time points (15 minutes post-facilitation and one week later). Assuming current clinical practice to be correct, multimodal 'orthographic and phonological' cue combinations were expected to deliver stronger facilitation outcomes, be it at the initial letter or whole word unit level. A secondary analysis was also pursued to explore the relationship between residual language abilities and the proportional naming gains from each cue condition. It was hypothesised that the amount participants benefit from each cue condition (if any) would vary, depending on their area(s) of language processing strengths/weaknesses.

Results from the group-level analyses suggested that a single application of either an 'Initial Letter: Orthography' or 'Whole Word: Orthography' cue resulted in facilitation 15 minutes later and a week afterwards. These effects were identified across the three outcome measures (Tables 3.3 to 3.5, or Table 3.6 for a summary). Effects from 'Whole Word: Orthography', in particular were robust, remaining significant on all three outcome measures, when compared against improvement arising from 'Repeated Exposure'. Group results for the other two cue conditions were not as reliable, with inconsistent and at best, marginal outcomes yielded across the outcomes. When tested against 'Repeated Exposure' where appropriate, none of the results from multimodal cues was significant, further suggesting that their effects might not be entirely attributed to the multimodal cues as their

improvements could not be properly distinguished from changes stemming from repeated encounters.

These set of results therefore provided direct support for the 'Orthography only' conditions, especially for 'Whole Word: Orthography' cues. They contradicted the prevailing clinical trend of applying multimodal cues wherever possible, but corroborated previous findings that the facilitation effects from one-off applications of written cues alone could last about 10 to 15 minutes (e.g., Best et al., 2002). These results are also novel on two fronts because previous studies either (1) did not assess effects of the whole word facilitation cue beyond the immediate 5 seconds (i.e., Howard & Harding, 1998) or (2) for studies which assessed effects 10 to 15 minutes post-facilitation, the researchers administered a different type of orthographic unit (e.g., Best and colleagues, 2002, used the CV written cue). In terms of quantifying naming gains that could be achieved, Best and colleagues (2002) provided a good basis of comparison, as theirs is the only study with a group-based analysis ($n = 11$; Table 3.1). They applied the CV written cue and elicited a mean naming improvement of 4.98 items (proportional change = 0.41), when assessed 10 minutes post-facilitation. In contrast, results in this chapter found that on average, a one-off application of the whole word written cue could improve final naming accuracy by 6.80 items (proportional change = 0.35) when assessed 15-minutes later (also see Appendix 3.5, which compiled the quantitative changes documented by other pieces of research, including single cases).

Crucially, the results presented in this chapter also provided new evidence extending the effects of written cues to a week. Previous facilitation studies involving a one-off exposure to an orthographic cue did not examine effects beyond around 15 to 20 minutes (e.g., Best et al., 2002). On average, the 'Whole Word: Orthography' increased the participants' final naming accuracy by 6.40 words (proportional change = 0.33) with a 95% confidence interval of [3.76 words, 9.04 words], and improved their naming latency by over 2200 ms with a 95% confidence interval of [420.35 ms, 4001.13 ms], when assessed a week later. These results are important as they provide fresh evidence for clinicians to plan for weekly sessions, should whole word orthographic cues be employed, with the confidence that for an average participant, there is likely a continuation of facilitation effects between sessions a week apart.

Of note, the significant effects found for the cue conditions were item-specific. No generalization to untreated items was detected, as inferred from the null findings consistently observed for the unseen controls (Tables 3.3 to 3.5). This is in line with findings on orthographic cueing reported in literature (e.g., Lorenz & Nickels, 2007). Clinically this also means that target words should be carefully chosen based on functional relevance, so that the items trained during a session could then be applied in daily life.

As for the language skills required to benefit from the cues (Secondary research question 4), the overall score calculated across assessments evaluating written language processing was associated with positive facilitation gains (higher naming accuracies and shorter latencies) across all four cues (results approached significance for initial letter multimodal cues; Table 3.7). When analysed using more specific language tasks, performance on word reading (CAT 20) stood out, as it was associated with positive facilitation gains across the cue conditions (again, results approached significance for initial letter multimodal cues). However to perform well on word reading, an individual could be engaging various residual language skills, be it lexical or sub-lexical. The data suggested that participants might be tapping on their lexical-semantic and/or sub-lexical skills, given the multiple marginally significant rank-order correlations observed between 'written synonym judgement' (lexical-semantic task) and different outcomes on 'Whole Word: Orthography' cues and 'Initial Letter: Orthography and Phonology' cues, as well as the marginally significant correlation observed between letter-sounding (sub-lexical task) and the RT outcome on 'Initial Letter: Orthography and Phonology' cue. These results meant that both the lexical (Lorenz & Nickels, 2007) or sub-lexical account (Best et al., 2002) are plausible at this stage. In chapter 4, the language pathways utilised by each participant will be further explored.

An intriguing question is why participants did not appear to profit from the multimodal cues combining orthography and phonology. At the group level, there was no strong support for either of the multimodal cues (Table 3.6). This is somewhat puzzling, since the orthographic forms were also embedded within multimodal cues, yet they failed to facilitate the significant improvements generated by 'orthography only' cues. One reason could be that the phonological cue distracts participants, thus reducing the impact of orthography. Presenting just the written forms behoves the participants to actually interpret the cues (e.g., see the words and make an effort to either read silently or aloud). In contrast, presenting both orthographic and phonological cues do not result in the same depth of processing – Participants may simply perceive the spoken cues at the expense of engaging fully with the written cues. This account receives tentative support from the group-level correlations, which featured significant associations between nonword repetition and outcomes from both multimodal cues but running in the contrarian direction: The better an individual was in producing nonsense sounds without the demands of semantic engagement, the poorer his/her naming output (Table 3.7). Nevertheless, having only the written form could have forced the participants to pay exclusive attention to the written stimuli and process accordingly. There could thus be deeper engagement with the solitary visual modality,

resulting in durable priming and thereby, stronger facilitation. While the above account is merely speculative, it serves as a promising fodder for future research. Similarly, it is unclear whether a staggered presentation of a range of orthographic cues followed by the attendant phonological cue (e.g., the setup found in a 'keyphone' device, Bruce & Howard, 1987) would deliver stronger facilitation outcomes compared to a simultaneous presentation of the two cues, since an individual could focus on one cue modality at a time. This again is an empirical question that warrants future investigation.

In conclusion, group level results from this facilitation case series provided evidence for cues delivered exclusively in their orthographic formats. In particular, evidence for the written whole word cue was robust, with effects lasting for at least a week on average, after a single application. Group-level analyses however failed to provide strong support for either multimodal cues that combined orthography with phonology. One should thus be cautioned against the indiscriminate application of multimodal cues. In the next chapter, the data will be examined at the individual level for additional insights.

Chapter 4

Comparing the facilitation effects of orthographic cues with or without the phonological components – Individual-level participant analyses

4.1 Introduction

4.1.1 Role of individual-level analyses and the use of a language processing model

Findings in Chapter 3 are based on group averages. These are important to help understand what might happen when an ‘average’ individual receives a specific facilitation cue. However, real clients seldom fit textbook cases and differ from the ‘average participant’ extrapolated by large-scale randomised controlled trials (Greenhalgh, Howick, & Maskrey, 2014, pp. 1 and 3). Importantly, group averages mask idiosyncrasies because single ‘average’ values computed across participants are not able to mathematically convey the complexities involving individual differences (Nickels & Howard, 1995).

In particular, there is substantial heterogeneity among people with aphasia post-stroke, which could impact outcomes from facilitation cues. Part of this heterogeneity stems from individual differences in lesion sites and sizes (Døli, Helland, Helland, & Specht, 2021). An example of how facilitation outcomes from group averages may differ from individual results in participants with brain injuries is illustrated by Crofts and colleagues (2004). On average, their group of 10 participants with word-finding difficulty improved significantly on spoken picture naming after completing a phonological facilitation task (i.e., word repetition in presence of the picture) when assessed 10 minutes afterwards. Yet when the same data were analysed at the individual-level, only one participant demonstrated significant improvement and two participants’ improvements approached significance. The remaining seven did not improve significantly following the phonological facilitation task.

For this reason, it is useful to complement group analyses with individual analyses. Individual analyses provide participant-specific outcomes (e.g., exactly which participant benefitted and to what extent), as well as attendant details (e.g., the participant’s key demographic and medical background information). With case-level information, a clinician is in a stronger position to determine how compatible the reported data were with the characteristics of a target client, and thereby make more meaningful therapy decisions tailored for the client.

One of the primary findings from the group analyses in Chapter 3 was the strong evidence in favour of the ‘Whole Word: Orthography’ facilitation cue. This was the only cue that delivered significant benefits on final naming accuracy and RT beyond the outcomes for

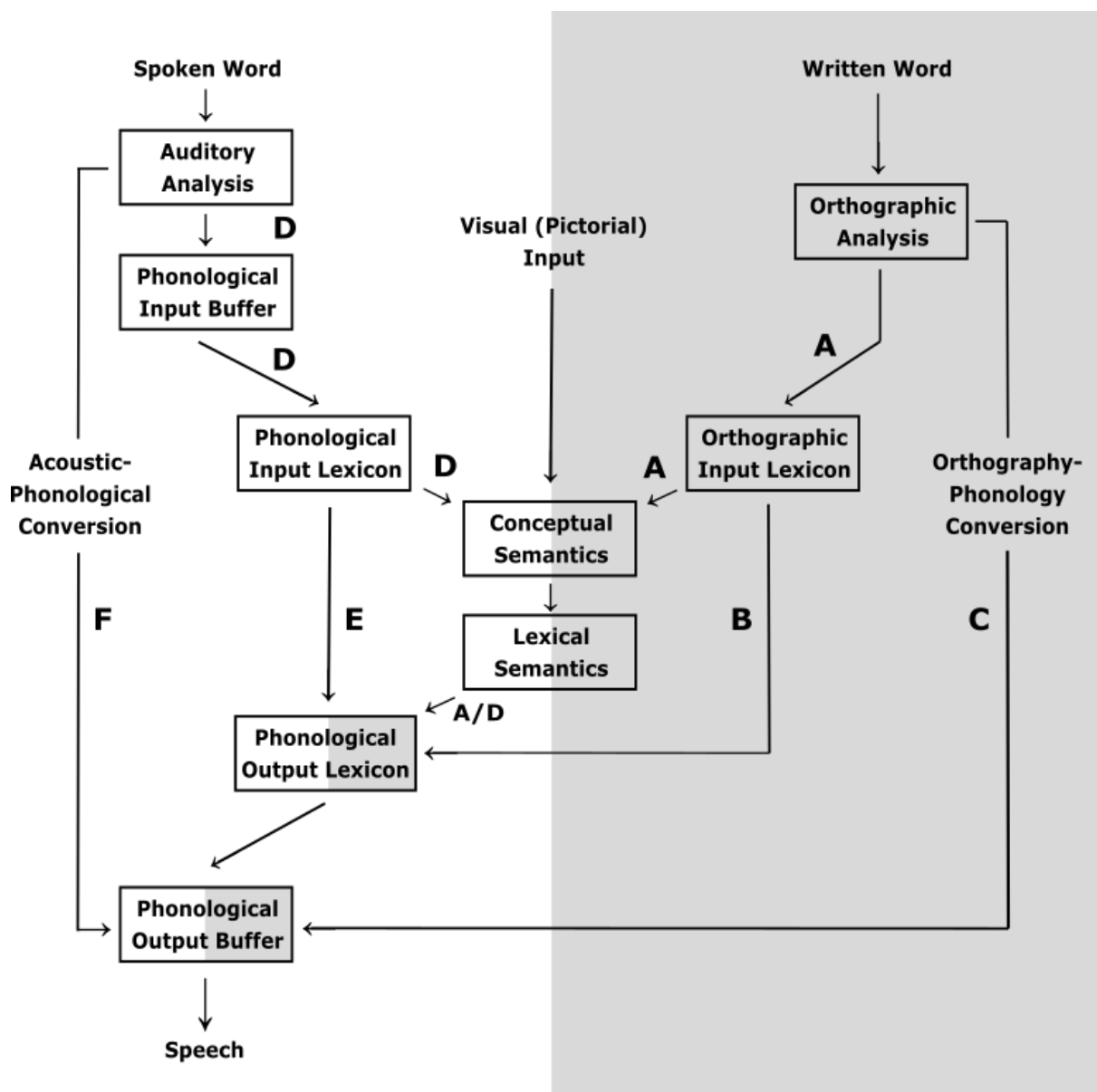
repeated exposure, when assessed both 15 minutes and one week later. Another noteworthy finding was that written language processing skills appeared to be correlated to beneficial facilitation outcomes for all four cue conditions (higher naming accuracies and shorter latencies). This could be because all of them contained an orthographic component. Out of the different written language processing tasks, reading aloud of words (CAT 20) was significantly correlated with beneficial outcomes; and this was observed on multiple measures (naming accuracies and/or RT) across the facilitation cues. However, as mentioned in Chapter 3, reading real words could involve several pathways, be it sublexical or the lexical routes (elaborated below). It is of clinical interest to attempt to determine exactly which pathways were engaged by each participant using individual analyses.

Figure 4.1 will be referred to, when processing pathways are explored. It is a reproduction of the same model mentioned in Chapter 1 (Figure 1.2, p. 6), which is adapted from Kay, Coltheart and Lesser (1992) as well as Howard and Harding (1998). Crucially, Figure 4.1 focuses on written and spoken inputs leading to a spoken output. It is correspondingly, a model of two halves and rendered so within the figure – the grey half indicating processes leading from written input towards the spoken output, and the white half for any spoken input leading towards a spoken output. Consequently, ‘Orthography (only)’ cues will be processed by the components in grey, while any cues incorporating the auditory input (e.g., ‘Multimodal: Orthography and phonology’ cues) will also activate the areas in white. Apart from semantics, shared between the two halves are the processes related to spoken output. This comprises the phonological output lexicon, phonological output buffer and pathways directed towards them or linking them. As mentioned in Chapter 1, impairment in any of the components and pathways may result in impaired spoken naming.

Important for this chapter, will be the input pathways as they represent how the different facilitation cues are processed to stimulate naming improvements. For cue conditions incorporating auditory input (e.g., ‘Whole word: Orthography and phonology’ cues), the phonological information is first subjected to preliminary analysis (e.g., segmenting and assemblage of sounds heard) as represented by the ‘auditory processing’ and ‘phonological input buffer’ components. Three different pathways then follow: First, there is the lexical-semantic path (path ‘D’) which is made up of phonological input lexicon and two semantic stores (conceptual and lexical semantics). The phonological input lexicon stores all the possible words an individual has heard of. Lexical semantics refer to the meaning representations specific to the word form (e.g., linguistic information), whilst conceptual semantics refer to the general aspects of the referent’s meaning beyond linguistic information (e.g., an object’s relationship with other items). In contrast, the direct

lexical path (path 'E') bypasses the need for comprehension, and directly activates the phonological word form in the output lexicon; this path could also be used when repeating real words. Lastly, the sublexical path (path 'F') converts the raw acoustic stimuli into its phonological units for articulation, and this is particularly relevant when repeating nonwords.

Figure 4.1. A cognitive neuropsychological model of single-word naming (adapted from Howard and Harding, 1998; Kay, Coltheart, and Lesser, 1992), with a focus on orthographic and auditory language inputs (written and spoken words), ending with spoken output.



To process the written stimuli (e.g., 'Whole word: Orthography (only)' cues), there are three analogous pathways. Similar to the auditory input processes, all written inputs are first subjected to a preliminary orthographic analysis. Three pathways for reading aloud are available to a skilled reader. The lexical-semantic path (path 'A') includes the orthographic input lexicon and the two semantic stores described above. The orthographic input lexicon contains all the written words an individual has encountered. The direct lexical path (path 'B') links the written input lexicon and spoken output lexicon together, thus bypassing the need to understand what is read. Finally, the sublexical path (path 'C') converts the individual graphemes into their corresponding sounds for spoken production. This last path is instrumental when sounding out letters and nonwords, or when reading aloud novel words (i.e., real words not encountered previously). Irregular words will be "regularised" using this route.

As mentioned in Chapter 3, there is clinical interest in determining whether there are common paths activated by all participants who benefitted from a particular cue. Results from such an analysis can potentially generate insights on cueing mechanisms. All cues in the facilitation study contain the written form, so understanding how orthography is decoded is therefore very relevant. To recap, evidence for processing the written cues through the sublexical route (path 'C') is provided by Best and colleagues (2002), as they identified a significant correlation between the participants' performance on a nonword task involving sublexical (path 'C') and their facilitation naming gains following written cues. Evidence for the lexical routes (either path 'A' and/or 'B') is supplied by Lorenz and Nickels (2007), because all three participants in their study had poor nonword reading skills but yet could benefit from the initial letter cues. Howard and Harding (1998) claimed that the direct lexical route (path 'B') was essential for their participant SD, after ruling out both sublexical (path 'C') and lexical-semantic paths (path 'A'). Given these accounts, it therefore remains an open question what the cueing pathway is, and whether there is indeed one primary processing route activated by participants with aphasia, in order to benefit from written cues.

4.1.2 Four research questions

The research questions in this chapter will resemble those asked in Chapter 3, except that the focus here will be on the individuals. Specifically, the interest here will be on how each participant responded to the facilitation conditions and how an individual's unique baseline language profile could be linked to his/her outcomes from a particular facilitation cue. The following questions were thus posed.

Primary research questions (1) and (2): At the individual participant level,

- **Which unit of the orthographic cue (initial letter versus whole word) and**
- **Which mode of cueing (orthography only versus multiple modalities with orthography and phonology combined) will result in significant improvement on the three main naming outcomes (initial/final accuracies, RT)?**

Similar to chapter 3, these two research questions investigate what constitutes an optimal cue involving the written form (unit of cue, number of modalities). Given the lack of published research directly comparing the different cue combinations, the verdict is open (Table 3.1). A successful combination should improve naming outcomes for the greatest number of participants.

Primary research question (3): How durable are the facilitation effects of each combination of cue (unit x modality) for each participant?

Following chapter 3, a successful cue combination should ideally result in improvements that last beyond the immediate period of time the cue was administered (e.g., one-week post-facilitation). Outcomes from individual analyses based on the first three questions will form the basis to group the five participants, according to the cue combination(s) they significantly benefitted from. This is because after a stroke, individuals may possess different residual capabilities. Correspondingly, they stand to benefit from different unique cue combinations. The grouping ties in with the secondary exploratory research question stated below.

Secondary research question (4): What is the relationship between each participant's baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?

By relating the residual language capabilities of the participants to their facilitation outcomes, it may be possible to identify common language processing pathways (Figure 4.1) used by participants who benefitted from the same facilitation cue. The caveat is the results related to secondary research question (4) will be exploratory because the analyses are associative and not driven by experimental manipulations; also, the sample is small. Individually-reported results from a small sample, however, are better able to assist practitioners in their future case-by-case comparisons.

4.2 Method

4.2.1 Design, procedure and stimuli

The design of the facilitation study, recruitment process, baseline assessments, stimuli and the three outcome measures (initial and final naming accuracies, RT) are described in detail in Chapter 3 (Section 3.2.1, p. 68). In this chapter, instead of averaging data across participants, data were preserved at the individual level and analysed accordingly.

4.2.2 Participants and clinical observations

Table 4.1 summarises the five participants' performances on baseline assessments. Consistent with their self-reports, the participants struggled with word-finding based on their impaired performances on spoken picture naming (all scored 3 *SDs* below the normative means on CAT 17 and 19 subtests). As mentioned in Chapter 3, participants also reported they had no hearing or significant visual impairments, and these were corroborated by their family members. They also achieved maximum scores on 'PALPA 18: Letter discrimination – Mirror Reversal' and could accurately repeat spoken words played to them during the preliminary equipment check.

The clinical impressions for each participant are described below in turn, with reference to Table 4.1. Table 4.2 summarises the interpreted information according to processing pathways in Figure 4.1.

4.2.2.1 P1's clinical profile at baseline and probable locus/loci of impairment

In terms of written language processing, P1 was severely impaired on all reading assessments, performing at close to floor on several subtests (e.g., PALPA 22: Letter Sounding, CAT 23: Reading Nonwords). An imageability effect could be discerned on word reading (CAT 20: Reading Words - High-imageability words = 75% versus low-imageability words = 0), suggesting that P1's lexical semantics were likely impaired, yet despite this, the lexical-semantic route (path 'A' in Figure 4.1) was still relied on for reading. This could be because other routes, like the direct lexical pathway (path 'B') and sublexical pathway (path 'C') were also impaired. A mild deficit on path 'B' could be inferred from the mildly impaired performance on visual lexical decision (PALPA 25). Severe impairment on path 'C' could be

Table 4.1. Scores on the background language assessments.

Assessment	<i>n</i> of items	P1	P2	P3	P4	P5
Spoken language processing						
PALPA 23. Spoken letter sound-written letter matching	26	1 (12.5%) *** (terminated after 8 items)	24 (92.31%)	8 (30.77%) ***	11 (42.31%) ***	6 (23.08%) ***
PALPA 23. Spoken letter name-written letter matching	26	22 (84.62%) *	26 (100%)	18 (69.23%) ***	17 (65.38%) ***	14 (53.85%) ***
PALPA 5. Auditory lexical decision task	160	137 (85.63%) ***	129 (80.83%) ***	132 (82.5%) ***	130 (81.25%) ***	134 (83.75%) ***
CAT 14. Repetition of nonwords	5 (Max score: 10)	10 (100%)	4 (40%) *	5 (50%)	8 (80%)	8 (80%)
PALPA 8. Nonword repetition task [†]	30	8 (26.67%) ***	7 (23.33) ***	8 (26.67%) ***	17 (56.67%) *	17 (56.67%) *
Written language processing						
PALPA 18. Letter discrimination: Mirror reversal	36	36 (100%)	36 (100%)	36 (100%)	36 (100%)	36 (100%)
PALPA 22. Letter sounding	26	1 (3.84%) ***	22 (84.62%)	0 (0%) ***	4 (15.38%) ***	5 (19.23%) ***
PALPA 22. Letter naming	26	1 (3.84%) ***	26 (100%)	11 (42.31%) ***	10 (38.46%) ***	2 (7.69%) ***
PALPA 25. Visual lexical decision task	120	111 (92.5%) *	108 (90%) *	101 (84.17%) **	100 (83.33%) **	93 (77.5%) **

CAT 20. Reading words	24 (Max score: 48)	18 (37.5%) ***	28 (58.33%) ***	24 (50%) ***	0 (0%) ***	14 (29.17%) ***
CAT 21. Reading complex words	3 (Max score: 6)	0 (0%) ***	0 (0%) ***	0 (0%) ***	0 (0%) ***	0 (0%) ***
CAT 22. Reading function words	3 (Max score: 6)	0 (0%) ***	4 (66.67%)	4 (66.67%)	0 (0%) ***	2 (33.33%) **
CAT 23. Reading nonwords	5 (Max score: 10)	0 (0%) ***	2 (33.33%) **	0 (0%) ***	0 (0%) ***	0 (0%) ***
PALPA 36. Reading nonwords	24	0 (0%) *** (terminated after 12 items)	5 (3.33%) ***	0 (0%) ***	0 (0%) ***	0 (0%) ***

Semantic processing

CAT 2. Semantic memory (similar to Pyramid and Palm Trees: 3 pictures version)	10 (Max score: 10)	10 (100%)	10 (100%)	7 (70%) **	9 (90%)	8 (80%)
CAT 7. Comprehension of spoken words	15 (Max score: 30)	28 (93.33%)	29 (96.67%)	25 (83.33%)	26 (86.67%)	21 (70%) **
CAT 8. Comprehension of written words	15 (Max score: 30)	28 (93.33%)	27 (90%)	24 (80%) ***	24 (80%) ***	21 (70%) ***
PALPA 49. Auditory synonym judgement [†]	60	46 (76.67%) ***	47 (78.33%) ***	30 (50%) ***	43 (71.67%) ***	40 (66.67%) ***

PALPA 50. Written synonym judgement [†]	60	51 (85%) **	53 (88.33%) *	35 (58.33%) ***	44 (73.33%) ***	41 (68.33%) ***

Naming

CAT 17. Naming objects	24 (Max score: 48)	35 (72.91%) ***	28 (58.33%) ***	19 (39.58%) ***	4 (8.33%) ***	20 (41.67) ***
CAT 19. Naming actions	5 (Max score: 10)	2 (20%) ***	4 (40%) ***	0 (0%) ***	0 (0%) ***	2 (20%) ***
CAT 25. Writing picture names	5 (Max score: 21)	21 (100%)	11 (52.38%) **	0 (0%) ***	7 (33.33%) ***	10 (47.62%) **

Both raw scores and percentages (in parentheses) are presented in the table.

* Borderline / Mild impairment: Between 1.5 to 2 *SDs* below mean OR within 1 *SD* below the cut-off score (if cut-off is provided, e.g., for CAT). This estimate is conservative and broadly follows Palmer, Boone, Lesser, & Wohl (1998), which diagnoses 1.3 *SDs* below mean as borderline and 2 *SDs* below mean as impaired.

** Moderate impairment: Between 2 to 3 *SDs* below mean or the cut-off score

*** Severe impairment: More than 3 *SDs* below mean or the cut-off score

[†] The norms for PALPA 8 (Nonword repetition), PALPA 49 (Auditory synonym judgment) and PALPA 50 (Visual synonym judgment) were not published in the official PALPA battery.

For PALPA 8, there are no published norms available. However, as P1 to P3 scored < 30% correct, they were deemed severely impaired on the task. In the absence of norms, there was less clarity for P4. However, a performance < 60% correct was likely at least mildly impaired.

For PALPA 49, the norms (*n* = 10) reported by Nightingale (1993) were used; and to the author's knowledge, this is the only set of norms available for this test.

For PALPA 50, the norms (*n* = 21) reported by Nickels and Cole-Virtue (2004) were used. Nickels and Cole-Virtue's norming samples for PALPA 50 were twice larger than another set of norms for the same test reported by Nightingale (1993; *n* = 10), so likely more representative of adults' performance and therefore, chosen for use.

Table 4.2. Summary of participants' language pathways relied on to process written/spoken input, based on their baseline performances.

Participant	<u>Language pathways involving written input</u>			<u>Language pathways involving spoken input</u>		
	Lexical-semantic: Path 'A'	Direct lexical: Path 'B'	Sublexical: Path 'C'	Lexical-semantic: Path 'D'	Direct lexical: Path 'E'	Acoustic-Phono Conversion: Path 'F'
P1 (Main naming impairment: Lexical semantics)	↓ but likely relied on	Mildly ↓, able to rely on			↓ but likely relied on (performance on auditory lexical decision = strongest within this sample)	
P2 (Main naming impairment: Phono output buffer)	Mildly ↓, able to rely on	Mildly ↓, able to rely on	✓ (Letter-sounding)			?
P3 (Main naming impairment: Semantic processing)	↓ but likely relied on			↓ but likely relied on		
P4 (Main naming impairments: Lexical semantics, access to phono output processes)	↓ but likely relied on					↓, able to rely on (CAT 14)
P5 (Main naming impairments: Lexical semantics, access to phono output processes)	↓ but likely relied on					↓, able to rely on (CAT 14)

‘✓’: Intact

‘↓’: Impaired

‘?’: Unclear

directly inferred from P1's performance on letter sounding (PALPA 22: Scored '1' out of '26') and nonword reading (CAT 23 and PALPA 36: '0' on each). Reading aloud nonwords, in particular, proved very challenging for P1, so much so that the task was terminated after 12 items. When reading aloud nonwords, lexicalisation was observed (e.g., P1 read 'fish' for 'shid' and 'doggies' for 'doop'), again suggesting a reliance on alternative reading pathways to substitute the use of path 'C'.

For language tasks involving spoken input processing, there was a discrepancy between P1's spoken letter-sound (PALPA 23: Severely impaired) versus letter-name (PALPA 23: Mildly impaired) matching. This indicated an impaired ability to convert graphemes into phonemes (path 'C'). Nonword repetition (PALPA 8) was also poor, implying possible deficits in acoustic-to-phonological conversion (path 'F') and/or the phonological output buffer. In terms of auditory (real) word processing, P1 was impaired on auditory lexical decision (PALPA 5), suggesting his auditory input lexicon could be impaired (part of paths 'D' and 'E'). However, when compared to the rest of the sample, his performance on auditory lexical decision (accuracy $\approx 86\%$) was the best, which could have implications for his ability to process auditory cues relative to others (discussed later).

For semantic processing, P1's performance on conceptual semantics (CAT 2: Semantic Memory) were within normal limits. However, for lexical-semantics (within paths 'A' and 'D'), investigation using PALPA synonym judgement tasks (PALPA 49 and 50) suggested there could be deficits: P1's performance was poor (at least 2-3 SDs below mean, Cole-Virtue & Nickels, 2004) with imageability effects replicated across written (high-imageability words: 28/30, low-imageability words: 23/30) and auditory (high-imageability words: 25/30, low-imageability words: 21/30) versions of the synonym judgement task.

Turning to spoken naming, the majority of P1's errors²² on object naming were semantic errors (60% of errors (3 out of 5), e.g., 'comb' for 'brush', 'fruit' for 'pineapple'), with some phonological errors (20% of errors, e.g., /ʌt.ə.flʌɪ/ for /bʌt.ə.flʌɪ/). Performance appeared weaker on low-imageability and multi-syllabic words (11/12 and 8/12 correct for high-imageability and low-imageability words; 11/12 and 8/12 correct for monosyllabic and three-syllable words). These were in line with impairments at semantic and post-semantic processes (e.g., phonological output buffer).

Taken together, P1's primary deficits for spoken picture-naming appeared to be firstly, at the level of lexical semantics, which was moderately impaired (part of paths 'A' and 'D'). This explained the large proportion of semantic errors he made during picture-naming. Mild deficits at the phonological output buffer were also likely present, given the phonological

²² Error analyses on CAT 17 and/or CAT 19 included incorrect initial responses and null responses, but excluded delayed but correct naming responses.

errors produced and length effect observed during naming. In terms of other language processes, P1's ability to convert orthography (graphemes) into phonology (phonemes; path 'C') was severely impaired, as evidenced by his poor nonword reading. The poor performance on nonword repetition further suggested his acoustic-to-phonological conversion ability was compromised (path 'F').

4.2.2.2 P2's clinical profile at baseline and probable locus/loci of impairment

For tasks involving written input language processing, P2 was impaired on some (not all) reading assessments. She was within normal limits on letter reading and sounding tasks (PALPA 22), with the latter suggesting that her ability to convert graphemes into phonemes (path 'C') was intact. However, there was a noticeable length effect and preponderance of phonological errors when she read aloud real words (CAT 20: 50% accurate on 3-syllable words and 75% accurate on monosyllabic words; $\approx$ 71% were phonological errors, e.g., /kɔ:t/ for 'cause, /kɔ:z/'). This pattern of results suggested an impairment of the phonological output buffer, which could explain the length effect as well as her poor nonword reading performance (PALPA 36: $\approx$ 52% were phonological errors, e.g., /pentʃ/ for 'prench'). In addition, her performance on the visual lexical decision task was mildly reduced, with a slight tendency to accept nonwords as words (PALPA 25: Correct acceptance of high-frequency words = 30/30, correct acceptance of low-frequency words = 28/30, correct rejection of nonwords = 50/60). This suggests her lexical-semantic and direct lexical pathways (paths 'A' and 'B') were likely at least mildly impaired.

For language tests involving spoken input processing, P2's performances on both spoken letter-sound and letter-name to written-letter matching (PALPA 23) were intact. Performance on auditory lexical decision was however impaired and appeared to be the weakest within this sample, suggesting likely deficits in her auditory input lexicon (paths 'D' and 'E'). Nonword repetition (CAT 23 / PALPA 36) was also impaired, and this was likely in part to due to phonological output buffer deficits (see length effects in spoken naming and reading aloud; length effects could not be reliably distinguished in nonword repetition because P2 performed at floor). Consistent with this account, the majority of her errors for nonword repetition were phonological errors ($\approx$ 86% of errors (19 out of 22), e.g., /plont/ for 'plonth'). The impairment at the level of the phonological output buffer also means that given the data, it is not possible to conclusively determine whether her acoustic-to-phonological conversion ability is impaired per se (path 'F').

In terms of semantic processing, conceptual semantics (CAT 2: Semantic memory) and lexical-semantics (CAT 7 and 8: Comprehension of spoken and written words) appeared within normal limits. Although P2's performances on auditory and visual synonym

judgements (PALPA 49 and 50) were impaired, the differences between high and low-imageability words on the tasks were small (written: high-imageability words: 28/30, low-imageability words: 25/30; auditory: high-imageability words: 24/30, low-imageability words: 23/30; average difference between high and low-imageability words across modality: 2) and she did not make semantic errors in object naming (CAT 17), suggesting that overall, her semantic impairment was likely mild (part of paths 'A' and 'D').

Specific to spoken picture-naming, all eight errors on P2's object naming (CAT 17) were entirely phonological (e.g., /pi.ri.ment/ for 'pyramid' and /pai.nup/ for 'pineapple'). Importantly, a clear length effect was present for P2 (12/12 and 4/12 correct for monosyllabic versus three-syllable words respectively). Both phonological errors and a length effect pointed to a deficit at the phonological output buffer.

Considering the assessment performance collectively, P2's impairment at the level of lexical semantics was mild (part of paths 'A' and 'D'). Her primary deficits resided in the auditory input lexicon (part of paths 'D' and 'E') and phonological output buffer. In particular, an impaired phonological output buffer would explain the unambiguous length effect in spoken naming and reading aloud, as well as the high proportion of phonological errors observed across tasks.

4.2.2.3 P3's clinical profile at baseline and probable locus/loci of impairment

Except for the reading aloud of function words (CAT 22) which was within normal limits, P3 was either moderately or severely impaired on the remaining CAT reading subtests. An imageability effect was also present in her reading aloud of real words (CAT 20: High-imageability words = 83.33%, low-imageability words = 16.67%). Thus, similar to P1, notwithstanding impaired lexical semantics (see below), P3 still depended on the lexical-semantic route for reading (path 'A'; Figure 4.1). Again, this dependence could be due to impairments in alternative pathways, including P3's sublexical route (path 'C'). A severely impaired path 'C' could be inferred from her poor performance on letter-sounding (PALPA 22) and nonword reading (PALPA 36), on which she scored zero. In addition, P3's performance on visual lexical decision (PALPA 25) was moderately impaired, suggesting that her visual input lexicon was compromised. This in turn meant her lexical-semantic and direct lexical pathways (paths 'A' and 'B') were affected to some degree.

Pertaining to spoken input processing, P3's performances on both spoken letter-sound and letter-name to written-letter matching (PALPA 23) were severely impaired. Her nonword repetition (PALPA 8) was also impaired, revealing both lexicalisation (e.g., 'change' for 'prench'; 'fire' for 'funior') and phonological (e.g., /kriəls/ for /kriəlθ/ ('crealth'); /gæfi/ for /gæfik/ ('gaffik')) errors. The lexicalisations of nonwords could suggest a flawed reliance on

the auditory lexical-semantic route (path 'D') due to an impaired acoustic-to-phonological conversion processing route (path 'F'). The phonological errors, however, hinted at further impairments in the post-semantic stages. Finally, her performance on the auditory lexical decision task was also impaired (PALPA 5). This indicated that her auditory input lexicon was likely compromised (i.e., paths 'D' and 'E' were affected).

Turning to semantic processing, P3 was moderately impaired on the conceptual semantic task (CAT 2) and severely impaired on a lexical semantic test (CAT 8: Comprehension of written words). Her selection of semantic distractors as responses in the semantic memory test (e.g., 'chair' for 'bed' or 'eye' for 'mouth') was in line with an impaired conceptual semantic store. Her poor performances on both auditory and written synonym judgment tasks (3 SDs below mean on PALPA 49 and 50 respectively), driven by very low rates of correct rejections across item sets (auditory: correct rejections for high-imageability words – 2/15, correct rejections for low-imageability words – 5/15; written: correct rejections for high-imageability words – 5/15, correct rejections for low-imageability words – 2/15), pointed to a severe deficit in lexical semantics (part of paths 'A' and 'D').

As for spoken naming, the errors P3 made on object naming (CAT 17) were predominantly a mixture of semantic errors ($\approx 23\%$ of errors (3 out of 13), e.g., 'ladybird' for 'butterfly') and non-naming responses, like pantomime or sound effects ($\approx 31\%$ of errors (4 out of 13), e.g., imitated the movement of a slug for 'snail' or hummed musical sounds for 'saxophone'). She also made erroneous responses that could not be classified ($\approx 23\%$ of errors (3 out of 13), e.g., 'duh' for 'star'; 'fish' for 'aerial'). Of note, P3's pantomimes or sound effects were not always accurate portrayals of the intended targets (e.g., making slurping sounds for 'toothbrush'). The latter reinforced the notion that conceptual semantics for P3 was impaired.

Overall, specific to spoken picture-naming, P3's primary locus of deficit was most likely, the semantic level, which was moderate-severely impaired (part of paths 'A' and 'D'). As described, deficits appeared to be present at *both* conceptual semantic and lexical semantic layers. These would account for her impaired performances across a range of tasks from conceptual semantic task (CAT 2: Semantic memory) to the semantic errors made during naming. As for other language processes beyond picture-naming, P3's impairments appeared to be distributed, with the remaining linguistic pathways compromised to varying extents (e.g., paths 'C' and 'F' were impaired, based on poor letter sounding and weak performance on nonword repetition).

4.2.2.4 P4's clinical profile at baseline and probable locus/loci of impairment for spoken naming

In terms of written language processing, P4's performance on visual lexical decision was impaired (PALPA 25: Correct acceptance of high-frequency words = 27/30, correct acceptance of low-frequency words = 25/30, correct rejection of nonwords = 48/60). This indicated an impaired visual input lexicon (part of paths 'A' and 'B'). He was, however, severely impaired on all assessments involving the reading aloud of words, scoring 0% across the CAT subtests. It was thus not possible to discern imageability or length effects. Performances on letter-sounding (PALPA 22) and nonword reading performances (PALPA 36) were also poor (< 16% correct), suggesting that the participant's sublexical skills were also severely-impaired (path 'C'). Interestingly, he was able to demonstrate understanding for some of the written words, despite his inability to read them aloud. For instance, in 5 out of the 24 test items within the reading word subtest (CAT 20), P4 was able to accurately and spontaneously convey the meanings through pantomime (mimicked equestrian movements for 'horse') or the production of sound effects (e.g., hummed the wedding tune for 'cathedral'). These responses implied that firstly, he was likely relying on the lexical-semantic path (path 'A') to read those words, and thus their correct meanings conveyed. Secondly, there were probably deficits residing in P4's post-semantic processes, which prohibited him from producing the sounds that made up the written words he actually recognised and understood.

For spoken input processing, P4 was impaired on the auditory lexical decision task (PALPA 5: Correct acceptance of high-frequency words = 35/40, correct acceptance of low-frequency words = 26/40, correct rejection of nonwords = 69/80), with a frequency effect present. These pointed to an at least moderate level of impairment within the auditory input lexicon (part of paths 'D' and 'E'). Although he performed within normal limits on 'CAT 14: Repetition of nonwords,' he was impaired on 'PALPA 8: Nonword repetition task.' The latter, comprised of more items, suggested mild deficits were likely present in the acoustic-to-phonological conversion pathway (path 'F'). There was also no length effect detected from his nonword repetition data (One-syllable: 5/10; two-syllable: 6/10; three-syllable: 6/10). P4 was further tested using the repetition of [real] words (CAT 12). He performed within normal limits (93.75%). The lack of a length effect and intact word repetition eliminated phonological output buffer as the primary locus of P4's deficit(s), shoring up support that an impairment at the acoustic-to-phonological conversion pathway (path 'F') underlie his impaired nonword repetition performance. P4's performance on nonword repetition was the strongest within the sample, suggesting this ability to be a relative strength compared to the others.

As for semantic processing, P4's conceptual semantics appeared intact as he performed within normal limits when directly assessed on it (CAT 2: Semantic memory). His lexical semantics, however, was severely impaired. This could be inferred from his impaired

performance on written word comprehension (CAT 8) and PALPA synonym judgement tasks in both written and spoken versions (more than 3 SDs below the mean in each of these three tests). Clear imageability effect was also present in P4's synonym judgement data (Written version: high-imageability words: 28/30, low-imageability words: 16/30; auditory version: high-imageability words: 24/30, low-imageability words: 19/30).

Finally, P4's performance on the picture-naming was almost at floor (e.g., accuracy for object naming (CAT 17): 8.33%), suggesting his spoken word retrieval difficulties were severe. The errors were a mixture of semantic ($\approx 14\%$ of errors (3 out of 22), e.g., 'sun' for 'star'; 'drive' for 'caravan'), phonological errors ($\approx 9\%$ of errors (2 out of 22), e.g., /strI/ for 'ski'), though non-naming responses dominated, e.g., pantomime or sound effects ($\approx 45\%$ of errors (10 out of 22)). P4's pantomimes and/or sound effects mimicked were accurate depictions of the target items. For example, P4 successfully gestured the movement of cutting for 'knife' and made drumming sounds for 'drum', although he could not provide the name for either. These suggested that a substantial part of his impairment resided in the post-semantic processes.

Pooling P4's baseline information together, the primary deficits affecting his spoken picture-naming would reside in his lexical semantics and the access to phonological output processes (i.e., post-semantic processes) (see paths 'A'/'D' on access). This combination of deficits would explain the participant's semantic errors during naming, imageability effects on semantic assessments, as well as his inability to produce the word aloud despite demonstrating adequate understanding of the item to be named or when he tried to read aloud a printed word. Another line of evidence to support this was P4's naming idiosyncrasies observed later. During baseline naming for treatment items, P4 would at times recite the alphabet before articulating the actual words, which were not always accurate (e.g., 'a, b, a, b, a, b, c, d ... dog' for 'zebra'; 'a, b ... bear' for 'bear'). The recitation of the alphabet perhaps represented a self-cueing strategy to jumpstart a failed access to phonological output (paths 'A' and 'D'), whilst the inconsistent semantic errors were likely a consequence of impaired lexical semantics. There were also impairments in the other language pathways affecting processes beyond spoken picture-naming. These included deficits in the visual as well as auditory input lexicons (part of paths 'A', 'B', 'D' and 'E') based on the lexical decision tasks, as well as P4's impaired sublexical pathway (path 'C') based on his reduced ability to read aloud nonwords.

4.2.2.5 P5's clinical profile at baseline and probable locus/loci of impairment for spoken naming

For written language processing tasks, P5 was moderately impaired on visual lexical decision, due to his poor rejection of nonwords (PALPA 25: Correct acceptance of high-frequency words = 30/30, correct acceptance of low-frequency words = 28/30, correct rejection of nonwords = 35/60 (58.33%)). P5's performance on this task was the weakest within the sample, and his tendency to accept most letter strings as real words was indicative of an impaired visual input lexicon (part of paths 'A' and 'B'). In addition, P5 was either moderately or severely impaired on the CAT word reading subtests (at least 2 *SDs* from respective means). On 'CAT 20: Reading words', no frequency effect was observed (performance both high and low-frequency words = 4/12) but a clear imageability effect was observed (high-imageability words = 7/12 (58.33%); low-imageability words = 1/12 (8.33%)). This indicated a likely reliance on the lexical-semantic pathway (path 'A') for word reading. His performance on letter-sounding (PALPA 22) was also poor, which meant his ability to convert graphemes to phonemes was likely impaired (path 'C'). This in turn contributed to his poor nonword reading (PALPA 36) – out of the 24 nonword items, he only produced single phonemic responses for 20 of them. Only 1 out of the 20 single responses was a correct initial letter sound (/d/ for 'doop'); the remainder bore no relation to the target sounds (e.g., /w/ for 'ked'; /v/ for 'smode').

As for spoken input processing skills, P5 was impaired on spoken letter-sound and letter-name to written-letter matching tasks (PALPA 23). He was also impaired on auditory lexical decision (PALPA 5), suggesting there were some deficits in his auditory input lexicon (part of paths 'D' and 'E'). Nonword repetition (PALPA 8) was mildly impaired, but the data did not reveal any length effect (one-syllable: 5/10; two-syllable: 7/10; three-syllable: 5/10). When tested using the repetition of [real] words (CAT 12), P5 scored full marks. These findings ruled out phonological output buffer as the primary locus of P5's deficit(s). The latter also reinforced the likelihood of a mild impairment at the acoustic-to-phonological conversion pathway (path 'F') to explain his nonword repetition performance. P5 was the co-high scorer with P4 on nonword repetition within the sample, suggesting this skill was also a relative strength of his, compared to others.

In terms of semantic processing, P5 performed within normal limits on the 'CAT 2: Semantic memory' subtest, suggesting his conceptual semantics was intact. However, his lexical semantics appeared to be severely impaired on both written and spoken modalities. His performances on both comprehension of spoken and written words tests (CAT 7 and 8) were at least two standard deviations below the norm, as were his performances on both PALPA written and spoken synonym judgment tasks. Moreover, imageability effects were present in P5's performances on synonym judgment across modalities (Written: high-imageability words: 25/30, low-imageability words: 16/30; auditory: high-imageability words: 22/30, low-imageability words: 18/30).

Turning to spoken picture-naming, P5's naming errors included semantic errors ($\approx$ 18% of errors (2 out of 11), e.g., 'frog' for 'snail') but the main bulk were null or incorrect verbal responses that were supplemented by attempts at finger-writing ($\approx$ 36% of errors (4 out of 11): In two of the errors, there were no spoken responses but correct finger-writing answers were produced, e.g., P5 said nothing yet finger-wrote 'BRUSH' for 'brush'; the other two errors were incorrect spoken responses but with correct finger-writing answers, e.g., said 'dog' but finger-wrote 'fox' for 'fox'). It should be pointed out that despite P5's attempts at finger-writing during spoken naming, when assessed directly on written naming (CAT 25: Writing picture names), his performance was moderately-impaired.

Overall, P5's primary sites of deficits were the lexical semantics as well as access to phonological output processes (i.e., post-semantic processes) (see paths 'A' and 'D'). Deficits in these areas would explain his semantic errors during picture-naming and imageability effects on semantic assessments, whilst P5's impaired post-semantic processes were manifested in his inability to verbally name some of the pictures, despite successfully spelling out the entire target word with finger-writing. In addition, P5 struggled with letter-sounding and reading of nonwords, suggesting a severely impaired ability to convert graphemes into phonemes (path 'C'). Finally, P5's performances on both visual and auditory lexical decision were impaired (part of paths 'D' and 'E' impacted based on latter); however, his score for visual lexical decision stood out for being the weakest amongst the five participants, suggesting a more prominently impaired orthographic input lexicon relative to other participants (part of paths 'A' and 'B'). Although an impaired orthographic input lexicon might not affect spoken naming directly, it will be interesting how that could impact his ability to profit from written cues.

4.2.3 Individual-level analyses for primary research questions (1) to (3): Outcomes of cue conditions across and between time points

Individual analyses were conducted at the item-level using WEighted STatistics (abbreviated as WEST by Howard, Best, and Nickels, 2015). The purpose of these item-analyses was to evaluate the overall efficacy of a particular condition across the four time points (i.e., Research questions (1) and (2)). To recap, the six conditions were:

- Condition (1): Initial Letter Cue: Orthography (e.g., 'a')
- Condition (2): Whole Word Cue: Orthography (e.g., 'apple')
- Condition (3): Initial Letter Cue: Orthography and Phonology (e.g., 'a' + /æ/)
- Condition (4): Whole Word Cue: Orthography and Phonology (e.g., 'apple' + /æpəl/)

- Condition (5): Repeated Exposure (i.e., Items not cued but encountered during facilitation), and
- Condition (6): Unseen Control (i.e., Items only encountered during assessments but not during facilitation)

WEST is a subset of the established contrast analysis technique (see Leustek, 2017), and has been well-applied to investigate small-*n* case studies in aphasiology (e.g., Barr, Biedermann, Tainturier, Kohnen, & Nickels, 2020). In particular, WEST tracks item-level progress over time. This is achieved by exploiting the sum of weights approach within contrast analysis to derive a summary score per item. In doing so, WEST crucially avoids having to input multiple scores for the same item measured at different time points, which would otherwise incur the risk of autocorrelation. Details on how to select weights and the exact steps taken for WEST are found in Appendix 5 of Howard and colleagues (2015).

Separate WEST-Trend and Rate of Change (ROC) analyses were conducted for each of the six conditions. To do so, two sets of summarised weighted scores were computed: (1) WEST-Trend, which determined whether there was a linear trend of improvement, and (2) WEST-ROC, which determined whether the change in performance was greater during facilitation/treatment phase(s) compared to non-facilitation/non-treatment phase(s). Given that the design consisted of two baselines and two post-facilitation data points, the corresponding weights assigned for WEST-Trend and ROC, for this chapter, were thus (-3, -1, 1, 3) and (1, -3, 3, -1) respectively. Finally, for each condition, the weighted results were then evaluated using a one-tailed one-sample *t*-test. Following Howard and colleague's (2015) recommendation, the manipulated condition should only be considered effective if the results were significant on *both* WEST-Trend and WEST-ROC.

While WEST examined whether a manipulated condition delivered a facilitation effect observed *across the entire duration* of the study, another important research goal was to clarify the exact duration of facilitation provided by each cue (i.e., Research Question (3)). In parallel to the steps taken for group analyses in Chapter 3, for the conditions found significant by WEST, two planned comparisons were run as follow-up using appropriate nonparametric tests determined by the properties of the outcome measure (i.e., McNemar for the binary accuracy outcomes, Wilcoxon signed ranks test for continuous RT). Each pairwise comparison would evaluate the performances (a) 15 minutes immediately post-facilitation and (b) one-week post-facilitation respectively, against baseline 2. The latter was chosen to index the participants' most recent performances prior to facilitation. For all

follow-up comparisons, Bonferroni's correction was applied to control for familywise error rates²³.

4.2.4 Individual-level exploratory analyses for secondary research question (4): Associations between performances on baseline assessments and facilitation outcomes

To explore the language mechanisms involved in the successful recruitment of a facilitation cue, one approach was to find out whether there were common language abilities shared by the individuals who profited from a particular cue. This was established using the following steps.

First, participants were grouped according to the facilitation cues that benefitted each of them. The groupings would be based on the results obtained from each individual's WEST and follow-up analyses. Specifically, significant naming improvements from the facilitation cue should be observed across time on both WEST-Trend and ROC, as well as beyond what could be achieved through mere repeated exposure.

Next, standardised z-scores were computed for each participant using their baseline language assessment scores. Consistent with the analyses in Chapter 3, these scores were divided into those indicating general language processes and performances on specific language assessments. Three overall average z-scores for the respective general language processes (written language (input), spoken language (input), and semantic processing abilities) were already computed earlier for each participant (see section 3.2.4.6.1) and those three values will be re-used here. As a refresher, for the z-score for a specific language assessment, performance from all five participants were pooled together to derive a group mean and standard deviation for that assessment (Step (a), p. 80). Using these values, for every assessment, a z-score was then computed for each participant (Step (b), p. 80).

Lastly, for each general language process and specific language assessment, participants were sorted according to their z-scores in ascending order, with the group median representing the middle score. Similar to the analyses in Chapter 3, only assessments with a sufficient item set size (i.e., smallest $n = 24$ items) would be used in the analysis of specific language assessment, for clearer discrimination of abilities between individuals (see section 3.2.4.6.2, or Figure 4.5 for the list of eligible assessments).

²³ Bonferroni's is applicable for item-level post hoc analyses in this study because unlike participant-level analyses, which relied on a small sample size ($n = 5$), the number of items for each item-level individual analysis is larger (e.g., ≈ 20).

4.3 Results

4.3.1 Primary research questions (1) to (3): At the individual participant level, which unit and mode of cueing will result in significant improvement; also, how durable are the facilitation effects of each combination of cue (unit x modality) for each participant?

Individual-level descriptive statistics for each of the three outcome measures are presented in Figures 4.2 to 4.4. Table 4.3 summarises the group and individual results, based on WEST²⁴ and the post hoc tests using McNemar or Wilcoxon, where appropriate. WEST analyses evaluated outcomes across all four time points. Post hoc comparisons, with Bonferroni correction applied, evaluated outcomes from the two post-facilitation time points (15 minutes or one-week later) against the final baseline (i.e., baseline 2) respectively. To ease reading, significant results in Table 4.3 are highlighted in yellow. Results for unseen (untreated) control items are not presented in the table as the outcomes were not significant across participants. Full statistical results for each participant can be found in the appendices. Appendix 4.1 contains the complete WEST-Trend and ROC results, including the null results for unseen controls, while Appendix 4.2 provides the detailed statistical output from all follow-up tests carried out when suitable.

The main findings for each facilitation condition are described as follow.

4.3.1.1 'Initial Letter: Orthography' facilitation cue

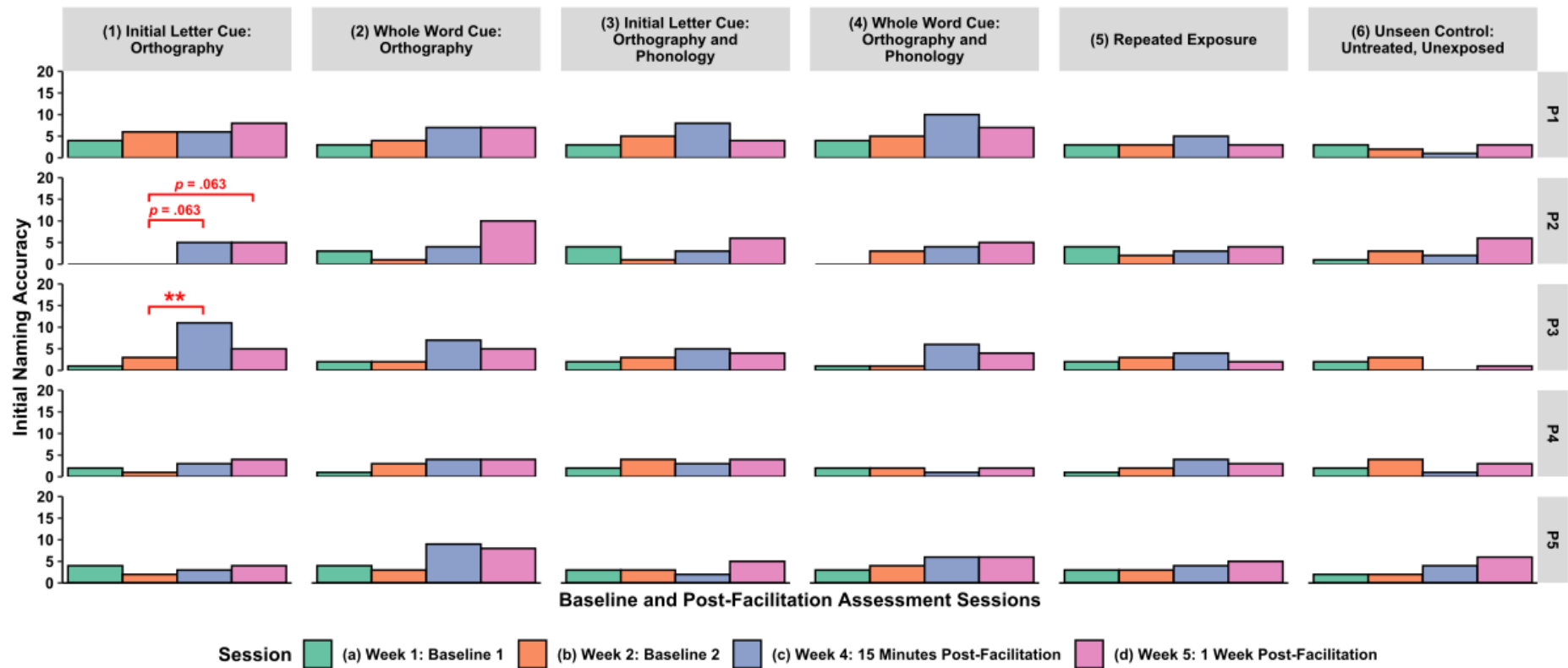
WEST-Trend and ROC analyses revealed that two participants (P2 and P3) demonstrated significant naming improvement, following the initial letter written cue. This was observed on initial naming accuracy for P2, and on all three outcome measures for P3 (Table 4.3).

For P2, neither of her post hoc results, based on initial naming accuracy, were significant after Bonferroni correction [both $ps = .0625$]. For P3, significant improvement over 'baseline 2' was identified when assessed on all three outcome measures 15 minutes post-facilitation [$ps < .008$ at least], but not one week later [$p > .20$ on each of the outcome measures; Appendix 4.2].

²⁴ WEST-ROC is a punishing test, as admitted in Howard et al. (2015), especially when item set sizes are small. It also penalizes participants who show an uptick in improvement on the final post-treatment test (e.g., P1's data on 'Whole Word: Orthography' condition, as his performance did not abide by the usual improvement trend). In WEST, a trend of improvement should show a performance that peaks shortly after treatment but declines as time wears on. To avoid Type-II errors, we interpreted conditions whose WEST results achieved marginal significance as well, i.e., $.05 < p \leq .07$.

Figure 4.2.

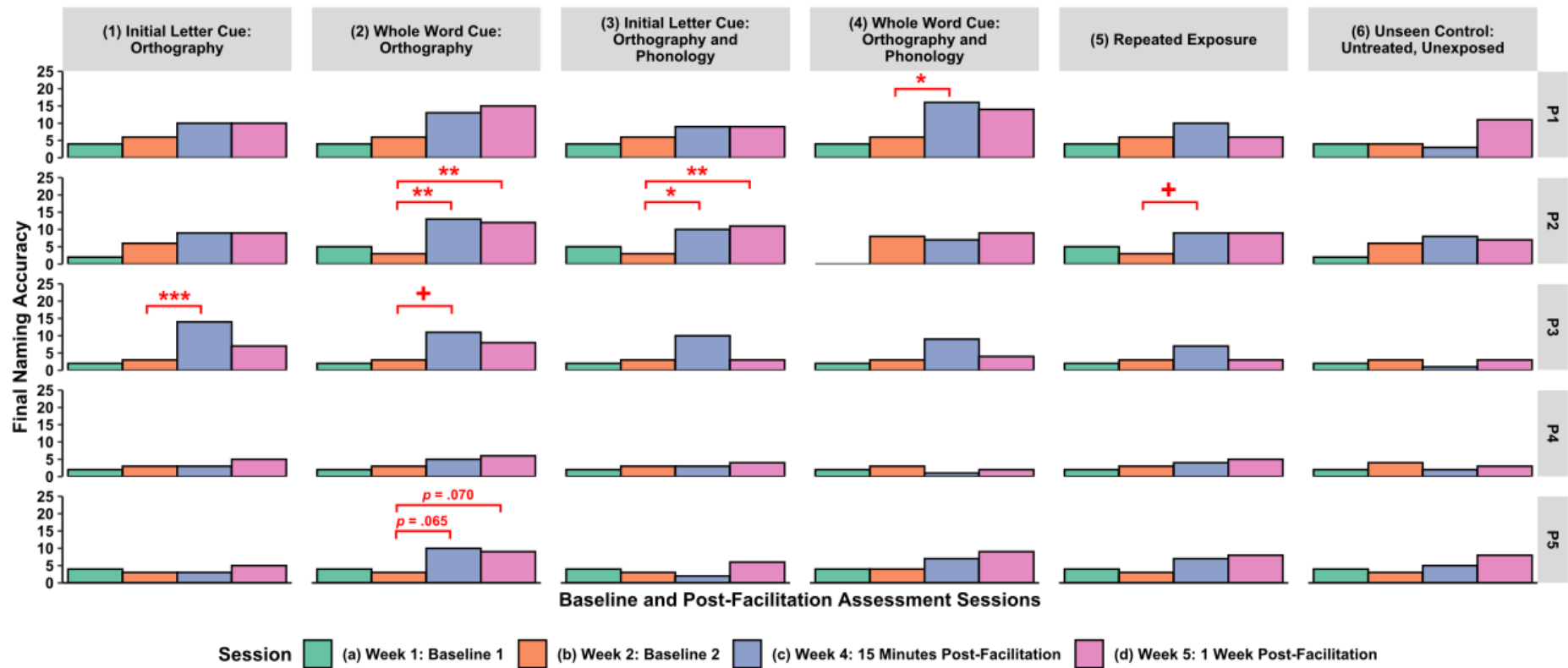
Initial naming accuracy scored for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively.



For post hoc: $+ .025 < p \leq .05$, $* .01 < p \leq .025$, $** .001 < p \leq .01$, $*** p \leq .001$

Figure 4.3.

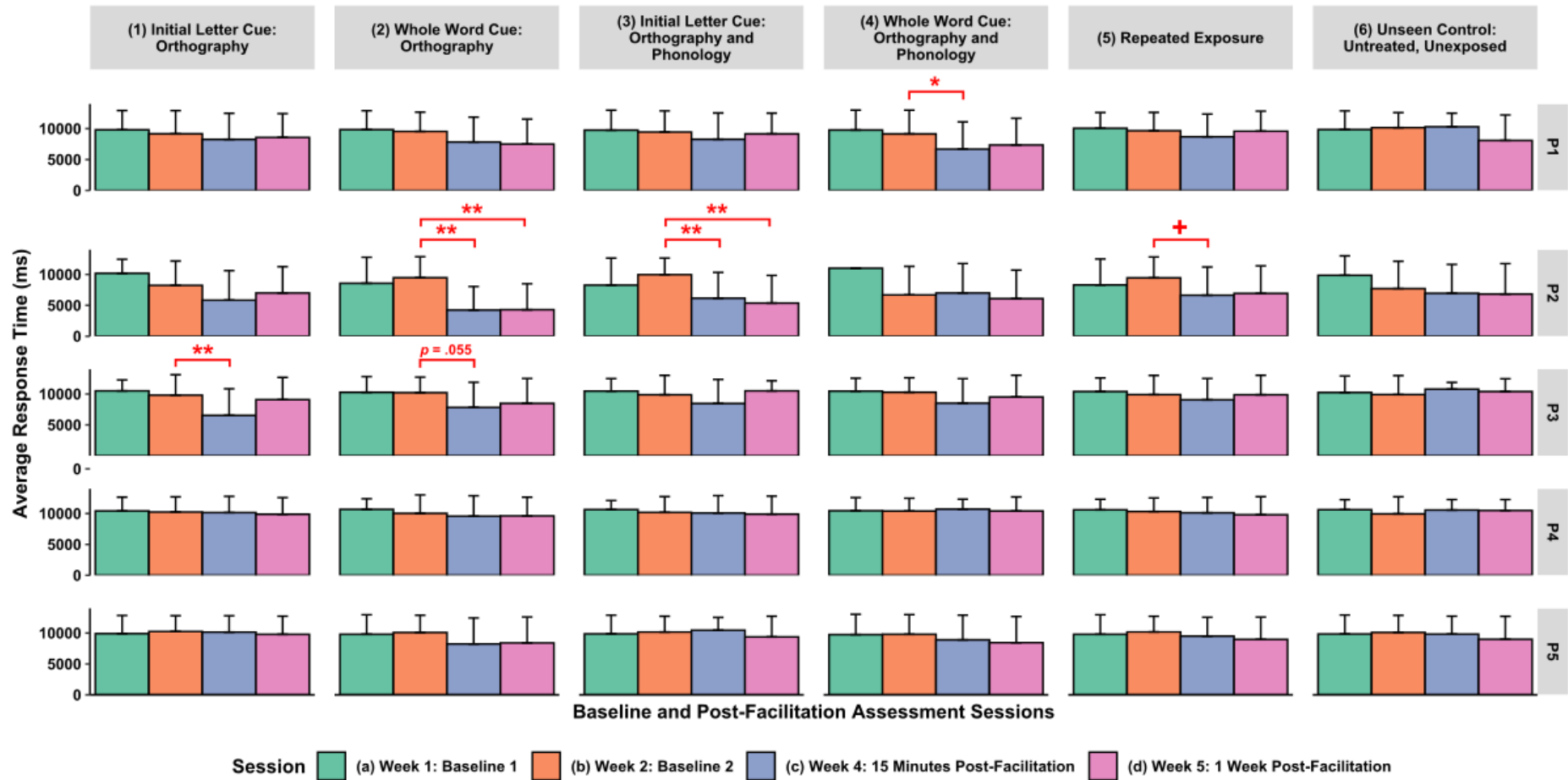
Final naming accuracy scored for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively.



For post hoc: + .025 < p ≤ .05, * .01 < p ≤ .025, ** .001 < p ≤ .01, *** p ≤ .001

Figure 4.4.

Average RT of correct responses for each participant across the four assessment sessions (two baselines and two post-facilitation) in the facilitation cues, repeated exposure and control condition respectively.



For post hoc: + .025 < $p \leq .05$, * .01 < $p \leq .025$, ** .001 < $p \leq .01$, *** $p \leq .001$

Table 4.3. Summary of facilitation findings for each condition for group and individual-level (WEST and nonparametric post-hoc) analyses for each main outcome measure.

Condition		Initial Letter Cue: Orthography			Whole Word Cue: Orthography			Initial Letter Cue: Orthography and Phonology			Whole Word Cue: Orthography and Phonology			Repeated Exposure		
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT
Group (Analysis by Participants)	Group results reproduced from Table 3.6 (p. 88)	*	*	*	**	**	**	<i>n.s.</i>	†	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	*
Post-hoc (against Baseline 2)	15 mins Post- vs. BL2 x1 Week Post- vs. BL2	† *	<i>n.s.</i> *	* *	* *	* *	*	–	<i>n.s.</i> †	–	<i>n.s.</i> †	–	–	–	* <i>n.s.</i>	* *
Post-hoc (against rep. exposure)	Final Baselines 15 mins Post- x1 Week Post-	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> † *	<i>n.s.</i> * *	<i>n.s.</i> * *	– <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	– <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	– <i>n.s.</i> <i>n.s.</i>	– <i>n.s.</i> <i>n.s.</i>			
P1 (W-Trend)		<i>n.s.</i>	*	*	†	***	**	<i>n.s.</i>	*	<i>n.s.</i>	†	***	**	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
P1 (W-ROC)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against Baseline 2)	15 mins Post- vs. BL2 x1 Week Post- vs. BL2	–	–	–	–	–	–	–	–	–	–	*	*	–	–	–
												<i>n.s.</i> (<i>p</i> = .077)	<i>n.s.</i>			
P2 (W-Trend)		**	**	**	**	***	***	<i>n.s.</i>	**	**	*	**	***	<i>n.s.</i>	**	†
P2 (W-ROC)		*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	**	**	<i>n.s.</i>	*	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	†	†
Post-hoc (against Baseline 2)	15 mins Post- vs. BL2 x1 Week Post- vs. BL2	<i>n.s.</i> (<i>p</i> = .063) <i>n.s.</i> (<i>p</i> = .063)	– –	– –	–	**	**	–	*	**	–	–	–	–	<i>n.s.</i> (<i>p</i> = .07) †	† <i>n.s.</i>
Post-hoc (against rep. exposure)	Final Baselines 15 mins Post- x1 Week Post-	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	– –	– –	–	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> *	– – <i>n.s.</i>	<i>n.s.</i> <i>n.s.</i> <i>n.s.</i>	– – <i>n.s.</i>	– – <i>n.s.</i>	– – <i>n.s.</i>	– – <i>n.s.</i>			

IN Acc.: Initial Naming Accuracy, FN Acc.: Final Naming Accuracy, Pts: Participants, W-Trend: WEST-Trend, W-ROC: WEST-ROC

BL: Baseline

WEST: † .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001 Follow-up tests (after adjusting for Bonferroni's): † .025 < *p* ≤ .05, * .01 < *p* ≤ .025, ** .001 < *p* ≤ .01, *** *p* ≤ .001

Comparing post-facilitation outcome against its repeated exposure counterpart (each time point compared just once): † .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001

–: Post hoc not required as effects of 'Repeated Exposure' were not significant, based on WEST for individual analyses

Condition		<u>Initial Letter Cue: Orthography</u>			<u>Whole Word Cue: Orthography</u>			<u>Initial Letter Cue: Orthography and Phonology</u>			<u>Whole Word Cue: Orthography and Phonology</u>			<u>Repeated Exposure</u>		
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT
P3 (W-Trend)		**	**	**	*	***	**	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	*	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
P3 (W-ROC)		**	***	**	<i>n.s.</i>	*	†	<i>n.s.</i>	*	*	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against Baseline 2)	15 mins Post- vs. BL	**	***	**	—	†	<i>n.s.</i> (<i>p</i> = .055)	—	—	—	—	<i>n.s.</i>	—	—	—	—
	x1 Week Post- vs. BL	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	<i>n.s.</i>	—	—	—	—
P4 (W-Trend)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*
P4 (W-ROC)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against Baseline 2)	15 mins Post- vs. BL	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	x1 Week Post- vs. BL	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P5 (W-Trend)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	**	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	*	**	**	*
P5 (W-ROC)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against Baseline 2)	15 mins Post- vs. BL	—	—	—	—	<i>n.s.</i> (<i>p</i> = .065)	—	—	—	—	—	—	—	—	—	—
	x1 Week Post- vs. BL	—	—	—	—	<i>n.s.</i> (<i>p</i> = .070)	—	—	—	—	—	—	—	—	—	—

IN Acc.: Initial Naming Accuracy, FN Acc.: Final Naming Accuracy, Pts: Participants, W-Trend: WEST-Trend, W-ROC: WEST-ROC

BL: Baseline

WEST: † .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001 Follow-up tests (after adjusting for Bonferroni's): † .025 < *p* ≤ .05, * .01 < *p* ≤ .025, ** .001 < *p* ≤ .01, *** *p* ≤ .001

Comparing post-facilitation outcome against its repeated exposure counterpart (each time point compared just once): † .05 < *p* ≤ .07, * .01 < *p* ≤ .05, ** .001 < *p* ≤ .01, *** *p* ≤ .001

—: Post hoc not required as effects of 'Repeated Exposure' were not significant, based on WEST for individual analyses

4.3.1.2 ‘Whole Word: Orthography’ facilitation cue

Three participants (P2, P3 and P5) significantly improved on their naming after receiving the whole word written cue, based on WEST-Trend and ROC. This was obtained on final naming accuracy and RT for P2 and P3, and just on final naming accuracy for P5 (Table 4.3).

In the post hoc analyses, significant improvement over ‘baseline 2’ performance was detected for P2 on both final naming accuracy and RT, at both 15 minutes and one-week post-facilitation [both final naming accuracy and RT: ps at least $< .004$]. Marginal improvement over ‘baseline 2’ performance was achieved by P3 at 15-minutes post-facilitation, when measured using final naming accuracy [$p = .039$]. The rest of the post hoc findings for P3 were not significant, after applying Bonferroni [ps at least $> .054$]. Despite demonstrating overall improvement over time on final naming accuracy, post hoc tests showed that P5 did not demonstrate significant facilitation effects at either post-facilitation time points [final naming accuracy: $ps \geq .065$; Appendix 4.2].

4.3.1.3 ‘Initial Letter: Orthography and Phonology’ facilitation cue

Based on WEST analyses, one participant (P2) improved her naming across time, when assessed on final naming accuracy and RT (Table 4.3). Post hoc analyses revealed that P2’s naming was significantly more accurate (final naming accuracy) and faster (RT) when assessed both 15 minutes and one-week post-facilitation [on both final naming accuracy and RT: ps at least $< .016$; Appendix 4.2].

4.3.1.4 ‘Whole Word: Orthography and Phonology’ facilitation cue

Two participants (P1 and P3) showed significant naming improvement, according to WEST-Trend and ROC (Table 4.3). This was demonstrated on final naming accuracy and RT for P1, and just final naming accuracy for P3.

In post hoc analyses, P1’s naming was significantly more accurate [final naming accuracy: $p = .013$] and faster [RT: $p = .025$] at 15 minutes post-facilitation; his results when assessed one week later was however not significant [final naming accuracy and RT: both $ps > .075$; Appendix 4.2]. For P3, in spite of her significant overall improvement on final naming accuracy across time based on WEST, her performance on the same accuracy measure did not significantly differ at either post-facilitation time points (15 minutes, one week later) from ‘baseline 2’, according to post hoc McNemar tests [final naming accuracy: both $ps > .10$; Appendix 4.2].

4.3.1.5 Distinguishing between improvement based on a facilitation cue versus spontaneous improvement based on repeated exposure

Out of the five participants, only one (P2) showed significant improvement, following repeated encounters with the pictures in the absence of cues, on two outcome measures (final naming accuracy, RT). Repeated exposure did not generate significant picture naming changes for the remaining four individuals, according to WEST (Table 4.3). For P2, additional tests were thus run, to ascertain whether her significant improvements post-facilitation, were reliably different from what could be achieved by mere exposure to the pictures while attempting to name them. Follow-up McNemar or Wilcoxon signed-ranks tests compared her facilitation outcomes that were significant based on WEST, against the corresponding performances she obtained under repeated exposure. It appeared that although P2 showed significant improvement following three types of facilitation cues ('Initial Letter: Orthography', 'Whole Word: Orthography', and 'Initial Letter: Orthography and Phonology'), only RT facilitation effects from 'Whole Word: Orthography' cue remained robust [$p = .046$], when compared against the effects from repeated exposure. Facilitation effects from 'Initial Letter: Orthography' and 'Initial Letter: Orthography and Phonology' did not reliably differ from her performance achieved with repeated exposure [Remaining comparisons: All $ps > .23$; Appendix 4.2].

To summarise, four groups of participants could be formed based on their WEST-Trend and ROC results. Specific only to P2, WEST analyses showed that repeated encounters with the pictures appeared to significantly enhance her naming outcomes. In view of that, for P2, only facilitation cues whose outcomes remained significant after comparing against analogous performances following repeated exposure, were taken as indicative of delivering facilitation benefits. The four groups were thus:

Group (1): Significant facilitation effects produced by the 'Initial Letter: Orthography only', 'Whole Word: Orthography only', and 'Whole Word: Orthography and Phonology' cues respectively (P3)

Group (2): Significant facilitation effects produced only by the 'Whole Word: Orthography only' cue (P2 and P5)

Group (3): Significant facilitation effects produced only by the 'Whole Word: Orthography and Phonology' cue (P1)

Group (4): No significant facilitation effect observed for any of the cues (P4)

The actual amount of improvement made by each participant in terms of raw values is of clinical interest. The information is compiled in Table 4.4. Only data whose post-facilitation outcome were significantly better than their respective performances achieved at 'baseline 2', are provided.

For participants who showed significant improvement throughout but whose data at specific time points did not significantly deviate from 'baseline 2' (e.g., P5 on the 'Whole Word: Orthography' cue), the exact *p*-values (these values happened to approach marginal significance)²⁵ were stated, where applicable. Also, in the case of P2, to mitigate confusion between actual facilitation effects from the spontaneous progress she made following repeated encounters, only facilitation outcomes that were significantly stronger than the performances achieved based on repeated exposure are reported. Based on the post hoc analyses of her data above, this meant only RT assessed one week after 'Whole Word: Orthography' cue was relevant.

For 'Initial Letter: Orthography' cue, the facilitation effect could be as high as 11 additional items successfully named in one's final responses (proportional change of 0.50), with the average naming speed lowered by 3266.9 ms (proportional change of 0.33), when assessed 15 minutes later (i.e., P3). 'Whole Word: Orthography' cue benefitted three individuals, and generated an improvement as high as 8 additional items (proportional change of 0.36) correctly named for final responses when assessed 15 minutes post-facilitation (P3), and naming speed enhanced by 5170.8 ms (proportional change of 0.55), when examined one week later (P2). Finally, for 'Whole Word: Orthography and Phonology' cue, when assessed 15 minutes afterwards, the largest improvement was registered by P1, which included 10 additional items successfully named in his final responses (proportional change of 0.42). His naming, tested at 15 minutes post-facilitation, was also on average quicker by 2464.4 ms (proportional change of 0.27).

4.3.2 Secondary research question (4): What is the relationship between each participant's baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?

Figure 4.5 arranged the participants according to their scores on each baseline task from low to high. In addition, each participant was colour-coded/grouped according to the facilitation cue(s) effective for him/her (see section 4.2.4 for the exact steps). The purpose

²⁵ Bonferroni was applied to post hoc testings against baseline 2, the pre-facilitation time point most immediately prior to facilitation cueing. Because of that, the criteria to attain significance or even to qualify as marginally significant was more stringent.

Table 4.4. Changes in the raw number of items correctly named and improvement in average RT (from baseline 2), as well as their proportional changes (Propn Δ). For P2, only post hoc results that remained significant after comparison against the corresponding repeated exposure data are reported here.

<u>Initial Letter: Orthography</u>							<u>Whole Word: Orthography</u>										
Change in Performance in:	<u>Initial Naming Accuracy</u>		<u>Final Naming Accuracy</u>		<u>Average RT (ms)</u>		<u>Initial Naming Accuracy</u>		<u>Final Naming Accuracy</u>		<u>Average RT (ms)</u>						
Time Post-Facilitation:	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week					
P1 (<i>n</i> = 30 per condition)											5170.8** Propn Δ: 0.55						
P2 (<i>n</i> = 16 per condition)																	
P3 (<i>n</i> = 25 per condition)													8** Propn Δ [†] : 0.36	11*** Propn Δ: 0.50	3266.9** Propn Δ: 0.33	8 [†] Propn Δ: 0.36	2380.0 (<i>p</i> = .055) Propn Δ: 0.23
P4 (<i>n</i> = 30 per condition)																	
P5 (<i>n</i> = 30 per condition)																7 (<i>p</i> = .065) Propn Δ: 0.26	6 (<i>p</i> = .0703) Propn Δ: 0.22
<u>Initial Letter: Orthography and Phonology</u>							<u>Whole Word: Orthography and Phonology</u>										
Change in Performance in:	<u>Initial Naming Accuracy</u>		<u>Final Naming Accuracy</u>		<u>Average RT (ms)</u>		<u>Initial Naming Accuracy</u>		<u>Final Naming Accuracy</u>		<u>Average RT (ms)</u>						
Time Post-Facilitation:	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week	15 mins	1 Week					
P1 (<i>n</i> = 30 per condition)									10* Propn Δ: 0.42		2464.4* Propn Δ: 0.27						
P2 (<i>n</i> = 16 per condition)																	
P3 (<i>n</i> = 25 per condition)									6 (<i>p</i> = .109) Propn Δ: 0.27								
P4 (<i>n</i> = 30 per condition)																	
P5 (<i>n</i> = 30 per condition)																	

For post-hoc (against baseline just prior to therapy, after adjusting for two comparisons using Bonferroni's): † .025 < *p* ≤ .05, * .01 < *p* ≤ .025, ** .001 < *p* ≤ .01, *** *p* ≤ .001

was to identify common language strengths shared by participants who benefitted from the same cue(s). Potentially, four viable patterns could be discerned.

- Pattern 1: This would distinguish individuals who benefitted from the 'Initial Letter: Orthography' cue from the rest, i.e., Group (1) from the rest. It involved identifying a baseline written language processing task in which the participant in Group (1) (P3) performed well, relative to the rest of the participants. To detect this pattern, the green bar (i.e., Group (1)) should be found on the far right. The remaining bars would be located to the left of it.
- Pattern 2: To distinguish individuals who benefitted exclusively from 'Whole Word: Orthography' cues (i.e., distinguish Group (2) from the other groups), the aim was to identify a *spoken language (input) processing task* in which the participants in Group (2) (P2 and P5) performed poorly in, compared to participants in Groups (1) and (3) (P3 and P1). This was because although participants in these three groups benefitted from cue conditions containing the whole word written form, only those in Group (2) did not benefit from any cue combinations containing a spoken component. To detect pattern 2, both orange bars (Group (2)) should be to the left of the green and blue bars (Groups (1) and (3) respectively).
- Pattern 3: To distinguish individuals who benefitted exclusively from 'Whole Word: Orthography and Phonology' (i.e., Group (3) from the rest), it would be a task that only participants in Groups (3) (P1) performed well in relative to the other groups. Thus, in pattern 3, the blue bar (Group (3)) should be positioned on the far right within the panel.
- Pattern 4: This pattern would help distinguish participants who *did not* benefit from any of the cue conditions involving written cues from individuals who did, i.e., Group (4) from the rest. The interest was thus to determine the baseline task(s) in which the participant in Group (4) (P4) performed badly, whilst others did not. To detect pattern 4, the aim was to look out for a task in which the purplish-pink bar was situated on the far left within the panel.

None of the panels in Figure 4.5 was in keeping with the descriptions spelt out for patterns 1 and 2. However, there was support for patterns 3 (boxed in blue) and 4 (boxed in red). For pattern 3, two of the panels were consistent with its description: (a) 'PALPA 5: Auditory Lexical Decision' and 'PALPA 25: Visual Lexical Decision'. P1 was the sole participant who only benefitted from the 'Whole Word: Orthography and Phonology' cues (Table 4.3). Although his performance on both lexical decision tasks were better than the

Figure 4.5. Participants' z-scores on baseline language assessments and their groupings according to the facilitation cue(s) found effective. Data sorted based on z-scores.



Table 4.5. Summary of facilitation cue(s) found effective for each participant and the language pathways they relied on to process written/spoken input, based on their baseline performances.

Participant	<u>Language pathways involving written input</u>			<u>Language pathways involving spoken input</u>		
	Lexical-semantic: Path 'A'	Direct lexical: Path 'B'	Sublexical: Path 'C'	Lexical-semantic: Path 'D'	Direct lexical: Path 'E'	Acoustic-Phono Conversion: Path 'F'
P1 (Main naming impairment: Lexical semantics)	↓ but likely relied on	Mild ↓, able to rely on			↓ but likely relied on (performance on auditory lexical decision = strongest within this sample)	
P2 (Main naming impairment: Phono output buffer)	Mild ↓, able to rely on	Mild ↓, able to rely on	✓ (Letter-sounding)			?
P3 (Main naming impairment: Semantic layers)	↓ but likely relied on			↓ but likely relied on		
P4 (Main naming impairments: Lexical semantics, access to phono output processes)						↓, able to rely on (CAT 14)
P5 (Main naming impairments: Lexical semantics, access to phono output processes)	↓ but likely relied on					↓, able to rely on (CAT 14)
'✓': Intact '↓': Impaired '?': Unclear Cues found to be effective: Initial Letter: Orthography only; Whole Word: Orthography only; Whole Word: Orthography and Phonology Whole Word: Orthography only Whole Word: Orthography and Phonology						

other participants, it was on 'PALPA 5: Auditory Lexical Decision' that this difference appeared salient, with his score approaching almost 1 *SD* higher than the next best participant's [Difference = 0.93 *SD*].

The distribution of participants in three panels corresponded with pattern 4's description. The three panels belonged to the following tasks/scores: (a) Overall average z-score across all written language processing tasks administered during assessment, (b) 'CAT 20: Word Reading', and (c) 'PALPA 50: Written Synonym Judgement' (see the three panels boxed within red rectangles). P4 was the only participant who did not significantly improve on any of the four facilitation cues, according to WEST (Table 4.3). On all three tasks/scores, he performed poorly compared to the other four individuals. This discrepancy between P4 (purple bar in Figure 4.5) and the rest was particularly prominent in 'CAT 20: Word Reading' – P4's z-score when reading aloud words was more 1.29 *SDs* lower than the next participant's, P5's, when sorted in ascending order. In comparison, P4's z-score was only less than 0.25 *SD* lower than the next participant's performance on the 'average z-score across all written language processing tasks' and 'PALPA 50' respectively.

Finally, the groupings of the participants are inlaid within Table 4.2, creating Table 4.5. By integrating information from the above clinical profiles (section 4.2.2 and table 4.2) and cues identified as effective for each participant, a clearer picture on how the locus/loci of impairment may affect outcomes from specific cues can emerge. This will be discussed below.

4.4 Discussion

To recap, group-level analyses in Chapter 3 revealed that both 'Initial Letter: Orthography' and 'Whole Word: Orthography' cues resulted in significant facilitation that persisted 15 minutes and one week later. These effects were replicated across all three outcome measures. In particular, facilitation effects from 'Whole Word: Orthography' cues remained significant when compared against effects from repeated exposure. In this chapter, these data were examined using individual analyses to uncover further insights.

Research questions (1) and (2): At the individual participant level, which unit (initial letter versus whole word) and mode of cueing (orthography only versus multiple modalities with orthography and phonology combined) will result in significant improvement?

Taking findings from WEST and post hoc comparisons against repeated exposure²⁶ into account, ‘Whole Word: Orthography’ cue emerged to be the most successful facilitation condition. It enhanced the naming outcomes of most participants, with three out of five individuals improving on at least one outcome measure (P2, P3, and P5). This was followed by ‘Whole Word: Orthography and Phonology’ cue (benefitted two participants, P1 and P3), and finally ‘Initial Letter: Orthography’, which facilitated all three naming outcomes for one participant (P3). In contrast, ‘Initial Letter: Orthography and Phonology’ cue did not facilitate the performance of any participant (Table 4.3 provides a summary, Appendices 4.1 and 4.2 detail the specific results). Although P4 did not benefit from any cues statistically, qualitatively, he performed best on the ‘Whole Word: Orthography only’ cue condition, achieving highest final naming accuracy and fastest RT following it, when assessed both 15 minutes and one week later.

It is worth noting the improved outcomes following ‘Whole Word: Orthography’ facilitation cues. Their effectiveness received support from both group and individual-level analyses, and the statistical benchmarks for each set of analysis were stringent: In group analyses, the facilitation outcomes were evaluated against performances based on repeated exposure, and for individual analyses, results had to be significant across WEST-Trend and ROC approaches. Given that the effects remained significant when subjected to this tough statistical scrutiny, there should be confidence in the efficacy of whole word written cues as facilitation prompts.

Individual analyses also yielded additional insights, as they preserved the idiosyncrasies between participants otherwise obscured by averaged outcomes in group analyses. According to group-level results in Chapter 3, aside from ‘Whole Word: Orthography’, on average, participants did not appear to benefit from other cue conditions. In contrast, individual analyses in this chapter revealed that some participants did significantly improve after receiving other facilitation cues (e.g., P3 with the initial letter written cue), suggesting that an individual’s facilitation outcomes could differ from the group average. Thus, for a more holistic picture, both group and individual analyses are necessary (see Best et al., 2019, for a more in-depth discussion on this point).

Lastly, individual-level analyses identified at least one participant, whose outcomes appeared to be significantly impacted by repeated exposure (i.e., P2). This was in line with

²⁶ Comparisons against repeated exposure applied only to P2. See section 4.3.1.5 for elaboration.

previous research, in which repeated exposure significantly influenced the naming performance of some but not necessarily all individuals (e.g., in Creet et al., 2019, repeated exposure improved naming of untreated items for 4 out of their 23 participants). However, previous research involved multiple encounters sustained throughout therapy. The results in this chapter thus represent an extension as they demonstrate potential improvement based on a *single* untreated encounter. In addition, the fact that not all participants were similarly impacted by repeated exposure as P2 was reinforces the idiosyncrasy in outcomes associated with people with anomia, which can be more reliably determined using a more sensitive approach like individual analyses. The results for repeated exposure are an important reminder for researchers to account for non-cue related improvements where applicable, to avoid spurious conclusions.

Research question (3): How durable are the facilitation effects of each combination of cue (unit x modality) for each participant?

Out of the four participants with significant gains, two demonstrated significant improvements when assessed 15-minutes after a single administration of one of these three cues, 'Initial Letter: Orthography' (P3); 'Whole Word: Orthography' (P3), and 'Whole Word: Orthography and Phonology' (P1). The 15-minute duration was approximately similar to what had been documented by others using facilitation cues involving written forms (see Appendix 3.5). For example, Best and colleagues (2002) reported facilitation effects lasting around 10 minutes for their participants, following a written CV cue.

Crucially, results obtained in this chapter extended the period of facilitation beyond the day the cue was administered. P2's RT data, following the 'Whole Word: Orthography' cue, demonstrated that facilitation gains could be maintained for at least a week. This raises the exciting possibility that for some people with anomia, effects of a one-off application of a written cue could potentially be more durable than previously thought. Practically, it could mean that for these individuals, fewer dosages and once-a-week sessions may improve spoken word-retrieval. Corroborating evidence from other participants in future research would be essential.

Secondary research question (4) and related clinical implications: What is the relationship between each participant's baseline performance on particular language assessment(s) and the ability to benefit from each combination of cue (unit x modality)?

By sorting the participants according to their baseline performances on standardised language assessments and associating these to their facilitation cue outcomes, two baseline tasks stood out: Each seemed useful in distinguishing (thus potentially predicting) either an individual who benefitted exclusively from 'Whole Word: Orthography and Phonology' cues (P1) or an individual who did not benefit from any of the manipulated written cues (P4). To distinguish the former type of individual, the task was 'PALPA 5: Auditory Lexical Decision,' and for the latter, 'CAT 20: Word Reading' (Figure 4.5). On both tasks, there was a difference of approximately 1 *SD* between P1 and P4 respectively from their next best performers, which would be a conservative benchmark.

These results suggested that firstly, for multimodal cues involving phonology, it could be sensible to screen suitability using an auditory lexical decision task. Secondly, as indicated by P4's baseline performance, to successfully capitalise on any facilitation cues manipulated in this study, the ability to read aloud the written form is likely to be important. This is perhaps not entirely surprising, given that each of these facilitation conditions contained a variation of the orthographic cue. Thirdly, no further systematic patterns could be reliably drawn (i.e., no results fitting patterns 1 and 2) which could be due to the heterogeneity in language abilities in people with anomia.

Finally, in terms of distinguishing which path was most likely recruited by people with aphasia to process a specific facilitation cue, results from this analysis could not pinpoint any common pathways. Rather, it appeared that the five participants had different loci of impairment (see section 4.2.2) and were relying on whichever residual path that functioned most effectively for them post-stroke. To have a firmer grasp on the matter of pathway(s) utilised, Table 4.5 is relevant as it lays out the pathways likely used by the participants to significantly benefit from each of the three cues.

Overall, Table 4.5 suggested that most participants depended on the lexical-semantic path (path 'A') for written cues, which was in line with Lorenz and Nickel's (2007) claim. However, data from participants like P2 means that the use of direct-lexical and sublexical pathways (paths 'B' and 'C') proposed by Howard and Harding (1998) and Best and colleagues (2002) respectively, cannot be dismissed either. Written cueing is therefore versatile in terms of operation – people with anomia do not approach an orthographic cue strictly from a single processing pathway, but can utilise at least one of the three potential pathways. Having more than one processing option, increases its likelihood of benefitting more individuals, depending on the language pathways unaffected by their neurological lesions (also see General Discussion, section 7.6).

Augmenting the evidence base in aphasiology at the case-level

Clinicians and researchers are often interested in the size of facilitation gains for various clinical purposes, including the gauging of clinical expectations. Thus, it will be useful to specify how the values obtained from case studies in this chapter relate to what has been reported in extant literature. To assist readers, the changes in raw values from published individual case evidence were extracted and compiled, to the best of the author's abilities, into Appendix 3.5.

For 'Whole word: Orthography' cues, case-level evidence from three participants (P2, P3, P5) are novel contributions to the literature. They demonstrated that a one-off application of the written word has the potential to significantly improve spoken word retrieval by increasing one's accuracy of final responses by 8 items (proportional change of 0.36) when assessed 15-minutes post-facilitation (as observed in P3), and 6 items a week later (proportional change of 0.22 in P5). It can also potentially speed up naming by 5170.8 ms (proportional change of 0.55 in P2; also see Table 4.4). These findings are important because Best and colleagues (2002; Study 4a in Appendix 3.5) have hitherto provided the closest evidence (mean change in raw number of items = 4.98, representing a proportional change of 0.41, when assessed 10 minutes post-facilitation), but their results were based on CV written prompts, not the whole word level.

Another novel contribution are the two pieces of case evidence on 'Whole Word: Orthography and Phonology' cues (P1 and P3). A single post-response administration of the whole word multimodal cue could significantly increase an individual's final naming accuracy by 10 items (proportional change of 0.42) with a reduction of naming latency by 2464.4 ms (proportional change of 0.27; both values from P1; Table 4.4), when assessed 15 minutes later. As far as this author understands, there is no facilitation data on this cue type published thus far. Lastly, previous evidence on 'Initial Letter: Orthography' cue was from Howard and Harding (1998), but they assessed only final naming accuracy at a very short time lag of 5 seconds post-facilitation. P3 extended the knowledge base by providing new case-level evidence on improvements related to different dimensions of naming when assessed at a much longer post-facilitation interval of 15 minutes (i.e., able to name 8 more items correctly (proportional change of 0.36) as the first response, 11 additional items (proportional change of 0.50) as the final response, and could be faster by 3266.9 ms (proportional change of 0.33; see Table 4.4).

Clinical implications

Echoing what was discussed in chapter 3, the results in this chapter cast further doubt on the prevailing clinical practice of administering multimodal cues for every client. It

is clear from individual analyses that not every participant in this sample benefitted from either 'Initial Letter: Orthography and Phonology' or 'Whole Word: Orthography and Phonology' cues. In fact, most did not, and no significant facilitation was observed with the initial letter multimodal prompts.

To ascertain the suitability of written cueing, screening clients by having them read aloud single words appear to be useful. This recommendation stems from the salient discrepancy detected on 'CAT 20: Word Reading' between individuals who benefitted from at least one cue combination involving the written form, versus P4 who did not (Figure 4.5). One reason why reading aloud words is an appropriate task would be that it is ecologically valid: The task assesses a crucial aspect of what individuals actually do during picture naming when the written cue is simultaneously provided, i.e., they often simply try to read or sound aloud the written form using whatever residual language skills they have. In terms of underlying mechanisms, to process visual word cues, participants are able to recruit at least one of the three written language processing paths (paths 'A' to 'C') outlined in Figure 4.1. Integrity of these three underlying pathways will similarly affect the performance on a single word reading aloud test. In other words, 'word reading' evaluates an individual's ability to negotiate his/her relative strengths and deficits embodied in paths 'A' and 'B', as well as potentially path 'C', when regular words are simply sounded aloud. As a corollary, the 'word reading' score best reflects an individual's eventual capability to profit from cues involving the written form.

To summarise, the facilitation study in Chapters 3 and 4 clarified the broad finding from the meta-analysis on the usefulness of orthographic cues. The efficacy was replicated across group and individual analyses, as well as in participants with different loci of impairments. The findings also extended previous research, including Best and colleagues (2002), by showing that even a single dose of orthographic cues can exert lasting outcomes, not just up to 10 minutes but a week later. In the next two chapters, the effects of using orthographic cues are further examined by way of a full therapy study.

Chapter 5

Comparing the use of initial letter against whole word as the unit of orthographic cue in spoken naming therapy – Group-level analyses

5.1 Introduction

Results from Chapters 3 and 4 demonstrated that orthographic cues were particularly effective when employed in isolation to facilitate spoken word retrieval. In these next two chapters, the rehabilitation consequences when written cues were applied multiple times, were examined. Of specific interest were (1) whether the orthographic cueing effects could last beyond one week, and (2) which of the two orthographic units, initial letter or whole word, would deliver stronger results.

There is a small pool of orthography-based *therapy* studies directed at improving spoken word-finding, and all of them were published before 2005. To recap, therapy is differentiated from facilitation, with cue therapy defined as the deployment of a cue through multiple applications sustained across time. This distinction between therapy (cumulative effect) and facilitation (a one-off cueing event) is in line with extant literature (e.g., Howard et al., 1985, p. 820). Table 5.1 summarises the relevant spoken anomia therapy studies involving orthographic stimuli published to date.

Although our interest exclusively centers on the use of orthographic cues, a few of these tabulated studies also compared the use of the written form against other modalities (e.g., phonological forms). In each, they demonstrated the superiority of a pure orthographic approach. Basso, Marangolo, Piras, and Galluzzi (2001), for instance, also examined the effects of phonological cueing (i.e., spoken repetition). Both their participants showed significant naming improvement following orthographic cueing, detected immediately or one to even five weeks post-therapy; however, spoken repetition only resulted in improvements immediately after treatment, but not at subsequent follow-ups. Others compared a semantic-based approach with an orthography-based therapy, and obtained similar results. For instance, P.A. did not improve on her naming after a semantic-based (word-to-picture matching) therapy designed by Nickels and Best (1996b). In Leonard, Rochon, and Laird (2004), M.H. did not produce notable improvement after the semantic features analysis therapy. Later, Leonard et al. also terminated M.H.'s phonological features analysis therapy, as "this treatment proved unsuccessful" (p. 185). Yet, both P.A. and M.H. demonstrated improved naming following their respective orthographic-based interventions. These discrepancies in therapy successes underscore the need to better understand how

Table 5.1. Summary of spoken word-finding therapy studies involving orthographic stimuli published to date. All studies focused on spoken picture naming accuracy as the primary outcome.

Study (Chronological)	Description of therapy	Main findings (post- vs. final baseline) based on the picture naming paradigm
Miceli, Armitrano, Capasso, and Caramazza (1996): Study 1 <i>n</i> = 1 (RBO: 38-year-old female; 18 months post-stroke)	Therapy technique: Pre-stimulation with written target words i.e., Read aloud all written words, then given picture to name Orthographic unit: Whole word Treated language: Italian Treatment frequency: 5 consecutive one-hour sessions	• Significant improvement on treated items observed immediately and 25 days post-therapy • No significant changes for untreated control items
Miceli et al. (1996): Study 2a: Therapy 1 (called 'Set 2' in paper) <i>n</i> = 1 (GMA: 60-year-old female; 12 months post-stroke)	Therapy technique: Pre-stimulation with written target words i.e., Read aloud all written words, then given picture to name Orthographic unit: Whole word Treated language: Italian Treatment frequency: 7 consecutive one-hour sessions	• Significant improvement on treated items observed immediately, 1 week, and 2 weeks post-therapy • No significant changes for untreated control items
Miceli et al. (1996): Study 2b: Therapy 2 (called 'Set 1' in paper) <i>n</i> = 1 (GMA: See above)	Therapy technique: Orthographic cueing - Written target words simultaneously presented with corresponding pictures Orthographic unit: Whole word Treated language: Italian Treatment frequency: 7 consecutive one-hour sessions	• Significant improvement on treated items observed immediately, 1, 2, 3 and 4 weeks post-therapy • No significant changes for untreated control items
Nickels and Best (1996b): Study 1: Therapy 1 (called 'letter cue therapy' in their paper) <i>n</i> = 1 (PA: 54-year-old female; 6 years post-stroke)	Therapy technique: Orthographic task – Participant was provided with a cue card with four letter choices printed on it ('s', 'f', 'd', 'b'), if unsuccessful in naming the picture. If still unsuccessful, participant read aloud the target word printed behind picture. Orthographic units: Initial letter and whole word Treated language: English Treatment frequency: 1 week, either 4 or 5 sessions	• Significant improvement on treated items observed immediately and 2 weeks post-therapy (see Set 1.1 on p. 126 in Nickels and Best, 1996b) • No significant changes for untreated control items when tested immediately post-therapy
Nickels and Best (1996b): Study 2: Therapy 2 ('lexical therapy' in their paper) <i>n</i> = 1 (PA: See above)	Therapy technique: Orthographic cueing – Participant read aloud the target word printed behind the picture, if unsuccessful in naming the picture. Orthographic unit: Whole word Treated language: English	• Significant improvement on treated items observed immediately, 1 and 2 weeks post-therapy (see Set 2.1 on p. 128 in Nickels and Best, 1996b)

	Treatment frequency: 1 week, either 4 or 5 sessions	• No significant changes observed for untreated control items when tested immediately post-therapy
Basso, Marangolo, Piras, and Galluzzi (2001): Therapy 1 (called ‘ Ortho Cue ’ in their paper) <i>n</i> =2 (RF: 66-year-old male; 6 months post-stroke; MR: 42-year-old male; 35 months post-stroke)	Therapy technique: Orthographic cueing – Progressive written cues provided if participant did not name the picture correctly within 5 seconds Orthographic unit: Initial letter progressing letter-by-letter to the entire word Treated language: Not stated Treatment frequency: 5 consecutive sessions	For both participants: • Significant improvement on treated items observed immediately and at various follow-ups (RF was assessed 1 and 12 weeks post-therapy; MR was assessed 1, 3, and 5 weeks post-therapy). • No significant changes for untreated control items
Basso et al. (2001): Therapy 2 (called ‘ Read Cue ’ or ‘ reading ’ in their paper) <i>n</i> =2 (RF and MR: see above)	Therapy technique: Orthographic cueing – Written whole word provided simultaneously with the picture for reading aloud Orthographic unit: Whole word Treated language: Not stated Treatment frequency: x5 consecutive sessions	For both participants: • Significant improvement on treated items observed immediately post-therapy, but no difference from baseline when assessed at various follow-ups (RF was assessed 1 and 12 weeks post-therapy; MR was assessed 1, 3, and 5 weeks post-therapy). • No significant changes for untreated control items
Hickin et al. (2002): <i>n</i> = 8 (Ages: 38 to 77 years)	Therapy technique: Orthographic task – Participants presented with a choice of written initial letter cues (e.g., ‘p’ or ‘k’), if their picture naming was unsuccessful. If still unsuccessful, the unit of written cues was increased to a syllable, and finally whole word. Orthographic units: Initial letter, syllable and whole word Treated language: English Treatment frequency: 8 weekly sessions (1 to 1.5 hours each)	• Results tested for overall therapy, but results specific to orthographic therapy were not separated out for testing. • Qualitatively, 7 out of 8 participants improved on treated items specific to orthographic task therapy, when assessed one-week post-therapy (Best, personal correspondence; e.g., HM: Final baseline = .44; Post-therapy = .60; PH: Final baseline = .38; Post-therapy = .68; also see Table 4 in Hickin et al.)
Leonard, Rochon, and Laird (2004): <i>n</i> = 1 (MH: 75-year-old male; 2 years post-stroke)	Therapy technique: Orthographic cueing – Progressive written cues provided or self-initiated, if unable to name picture correctly Orthographic unit: Initial letter progressing letter-by-letter to the entire word Treated language: English Treatment frequency: 14 sessions	• Only descriptive statistics available • Qualitatively, item-specific improvement observed immediately (likely) but not two weeks post-therapy (Baseline = 1/10; immediate (likely) post-therapy = 8/10; two weeks post-therapy = 1/10) • Qualitatively, no changes observed for untreated items (Baseline to two weeks post-therapy = 0/10)

orthography could be incorporated to rehabilitate word-finding.

It is also worth noting the diversity of therapy techniques identified under the umbrella of orthographic-based therapy (Table 5.1). A few points can be made: (1) Comparisons between these different orthographic techniques is beyond the scope of this study. Our focus will be on orthographic cueing, as it appears to be easily implemented, and thus most functional approach for direct application within the clinic. (2) Similar to the literature on facilitation studies (Chapter 2), there appears to be no standard unit of orthographic cue in the therapy studies. (3) All documented improvements from an orthographic cue appeared to be item-specific, with no significant generalisation to untreated items. (4) It appears that participants across age-groups and different timings post-stroke could benefit from orthographic-based anomia therapies.

A secondary interest is the cognitive underpinnings of successful therapy outcomes following orthographic cues. Understanding which cognitive variables are associated with therapeutic gains, will assist clinicians in determining the suitability of their clients for an orthographic cue-based therapy. This is in the context that individuals with post-stroke aphasia often display cognitive deficits (e.g., Murray, 2012). The typical breakdown of specific cognitive domains includes memory, attention, and executive function (e.g., cognitive flexibility); many of these domains are in turn tapped upon by a broader construct like general nonverbal reasoning (see p. 223 in Raven and Raven, 2003, whereby they described their nonverbal reasoning test as measuring “general cognitive ability”). There is a small but growing body of work in aphasia research investigating the cognitive predictors (specific and general factors) of positive therapy yields. Two studies are worth highlighting, given their larger sample sizes and proximity of their therapies to ours (both studies utilised multi-modal cueing including written cues, conducted within a spoken picture naming paradigm). Having bigger samples meant their findings were potentially more generalizable.

Lambon Ralph, Snell, Fillingham, Conroy and Sage (2010) combined data ($N = 33$) from their previously published studies on the basis that all their anomia therapies were broadly similar (specifically, they employed progressive multi-modal cueing whereby both written and spoken cues were simultaneously presented, albeit in either increasing or decreasing directions) and that the same cognitive assessments were administered at baseline. Results from their Pearson correlation analyses (p. 300) suggested that selective attention (elevator counting with distraction) and visuo-spatial memory (Rey-Osterrieth figure copying and delayed recall) were significantly related to naming gains at one-week and five-weeks post-therapy.

On the other hand, the COMPARE trial ($N = 140$; Rose et al., 2021) investigated the effects of a therapy like the Multi-modality Aphasia Therapy (in which one’s written

responses could function as a potential cue for spoken naming). Results from their multivariate regression revealed that general nonverbal reasoning skills (indexed by Raven's matrices) predicted naming gains for the participants with mild aphasia immediately post-therapy, whilst performance on an attentional task (elevator counting) did not. There is therefore a lack of clarity on exactly which cognitive ability is most critical for successful naming rehabilitation. If small-*n* findings were included, the evidence for any given cognitive factor becomes more uncertain (see Dignam et al., 2017, for a related summary). Importantly, to this author's knowledge, none of these published findings were specific to the therapy outcomes of a purely *orthographic-based* treatment.

In view of the above, the following research questions were formulated.

Primary research question (1): At the participant group level, which unit of the orthographic cue (initial letter versus whole word) will result in significant improvement on the three main naming outcomes (initial/final accuracies, RT), following multiple applications of each cue?

There are strong reasons in favour of either the initial letter or the whole word unit. By convention, prompts are initiated using smaller units. This means a smaller unit like initial letters are preferred, as they necessitate a level of effort from the participant, thereby generating deeper processing and a more durable memory trace in the long run. However, results from our facilitation study suggested that the whole word could potentially be as effective a cue as the initial letter. Anecdotally, the whole word unit is also the preferred unit by individuals with anomia (also see related research comparing increasing versus decreasing cues in the context of hierarchical multi-modal (spoken plus written) cueing, e.g., Fillingham and colleagues (2005, 2006), in which they found no difference in therapy outcomes between cue treatments providing initial letters first and those commencing with whole words; and that all their participants preferred receiving the entire word first). In the absence of data directly comparing the two orthographic units, the superiority of the initial letter or the whole word is an open empirical question.

Primary Research question (2): How durable are the therapy effects of each orthographic cue (initial letter versus whole word) across the participants?

To answer this primary research question, the performance trajectories of each written cue were tracked from one week (short-term effect) to at least four weeks (long-term effect) post-therapy across three main naming dimensions (initial and final naming

accuracies, response time). The participants and caregivers' views were also collected pre- and post-therapies using rating scales. Whichever the unit, the stronger orthographic cue would achieve significant improvement across the multiple outcome measures (naming accuracies, latencies, and/or participant ratings), with the positive changes sustained for longer periods of time.

Secondary research question (3): At the participant group level, what is the relationship between baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?

To answer this secondary research question, standardised cognitive assessments across the main cognitive domains (attention, memory, executive function, general nonverbal reasoning) were administered at baseline. Correlation analyses would be run between scores from baseline cognitive assessments and naming gains for each cue achieved at each post-therapy time point. There are a number of cognitive processes that may be necessary to benefit from orthographic cues: An individual needs to be able to attend to the prompt (attention), temporarily store the cue and retrieve the relevant spoken word form from the mental lexicon (memory), update one's working memory while discarding information irrelevant to the cue (executive function), and potentially 'problem solve' instances of word-retrieval failures by initiating a self-cueing strategy (general nonverbal reasoning). However, there is a lack of orthographic-based therapy studies to sufficiently inform which cognitive factor would be most important for each unit of written cue. Given our small sample size and the dearth of relevant literature, this investigation will be exploratory.

5. 2 Method

5.2.1 Design

A small-*n* case series approach was adopted, similar to the facilitation study (Chapters 3 and 4). However, to compare between the two cue therapies, an A-B crossover design was incorporated, with multiple baselines and a control task. Similar to the facilitation study, nonword repetition was chosen as the control task (PALPA 36, Kay et al., 1992) because the ability to listen and repeat spoken nonwords should not be affected by naming interventions, which entail orthographic cues.

Table 5.2. Participant information.

	Age	Gender	Highest level of education	Aetiology	Time post-onset (Years; months)	WAB score and diagnosis
P1	80	M	University	Left anterior MCA	14;11	84.4; Anomia
P2	75	F	Secondary School	Left MCA	8;7	69.6; Conduction aphasia
P3	55	F	Secondary School	Left-sided intracerebral haemorrhage	0;10	78.2; Anomia
P4	66	M	Secondary School	Left MCA	~ 3;0	66.6; Conduction aphasia
P5	70	F	University	Left total anterior circulation infarct	5;5	39.5; Broca's aphasia

5.2.2 Participants

To be included in the study, the participant should: (1) be an adult (i.e., at least 18 years old), (2) be a native speaker of British English, (3) have had a stroke that occurred at least 6 months prior to the start of the study, (4) have spoken word-finding difficulties post-stroke, and (5) have sufficient visual and auditory acuity to interact with a laptop computer. Exclusion criteria included participants with neurodegenerative diseases (e.g., dementia) and/or those with moderate or severe apraxia of speech. The latter was assessed using subtests 1, 2b and 3 from the Apraxia Battery for Adults (ABA-2, Dhabul, 2000; details in Appendix 5.1). From clinical experience, these three subtests or their variants are commonly used at bedside to assess apraxia of speech.

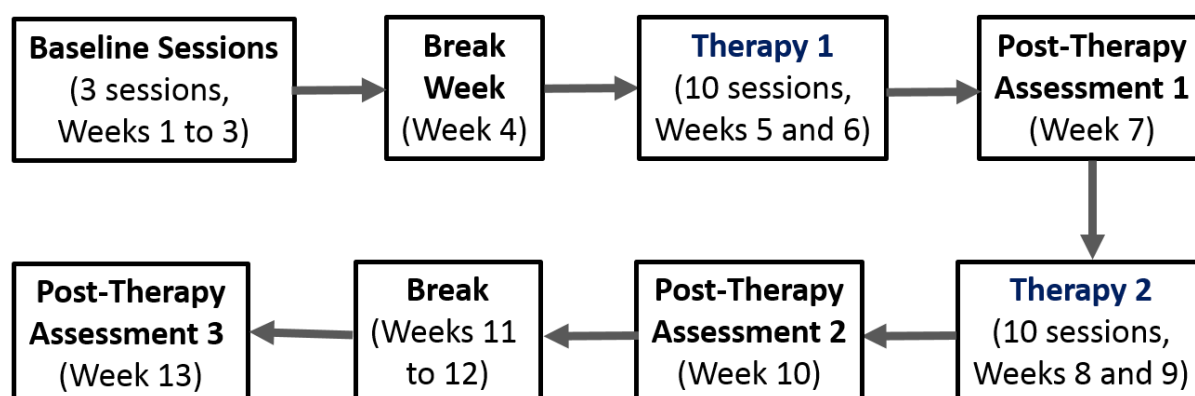
Recruitment took place over 10 months. Almost a hundred participants were contacted; these were potential participants listed on the university communication clinic register, direct referrals from therapists and aphasia organisations, or directly recruited in-person during public seminars and aphasia meetings. Out of these, thirty-one responded; seventeen eventually declined participation (e.g., change in personal circumstances), and one was not suitable as his aphasia was secondary to brain tumour, not a stroke. The remaining thirteen were evaluated face-to-face either in person (pre-Covid lockdown) or on Skype. Out of these, five had very mild anomia and performed at ceiling on picture naming (> 95% accuracy on the 327 stimuli pictures), and thus were not recruited. Three had severe or moderate-severe apraxia of speech, based on ABA-2 subtests, and were excluded. The remaining five respondents fulfilled all criteria and provided written consent to participate, making up the final sample.

The demographic and medical information for each participant is provided in the Table 5.2 (their individual performances on baseline assessments are published in Table 6.1). All five participants (2 males, 3 females; their ages ranged from 55 to 80 years old) were Caucasian, right-handed and had a left-sided stroke. Except P2 who was mostly home-bound due to her hemiplegia, the rest were community ambulant. None of the five participants reported major hearing and visual deficits, which were corroborated by their caregivers and all scored full marks on the visual letter discrimination test (PALPA 18: Mirror Reversal).

5.2.3 Procedure

Figure 5.1 shows the timeline of this study.

Figure 5.1. Flowchart of the therapy study.



5.2.3.1 Assessments (Baseline and post-therapy sessions)

The study spanned 13 weeks for each participant. Baseline sessions were held once a week, each lasting about two to two-and-a-half hours, with adequate breaks in between tasks (see section 3.2.4 for details on the breaks). An additional baseline session was scheduled on the third week for two participants, to accommodate their pace. Over the course of these baseline sessions, spoken picture naming for 327 coloured photographs were collected thrice. The photographs comprised the original 296 pictures used in the facilitation study (see Section 3.2.3 on p. 71), and 31 new items.²⁷ All pictures were digitally formatted to 500 by 500 pixels and 24-bits depth to maintain consistency and clarity. The baseline picture naming paradigm was identical to that of the facilitation study (see Figure 5.2a below), with no cues provided.

Post-therapy assessments 1 and 2 were designed to capture short-term effects of the therapy and were conducted one week after the completion of either Therapy 1 or 2. Post-therapy assessments 2 and 3 were included to examine long-term maintenance effects for Therapy 1 and 2 respectively (Figure 5.1). These took place on the fourth week after the completion of the respective therapy. During post-therapy assessments, the same picture

²⁷ 50 new photographs were initially sourced from online royalty-free photo repositories. A group of five healthy adult native speakers of English (3 males, 2 females; Age: Mean = 48 years; *SD* = 12.27) independently provided a written name for each picture. Only items fulfilling 80% naming agreement were chosen. 32 pictures fulfilled the criteria. These 32 pictures were then independently screened by another individual (a British-native speech therapist) to make sure all items were familiar to the local population, especially the elderly. One of the items, 'dishwasher', was dropped in the process, as this appliance is not found in all UK households. The same therapist also verified that all target words began with an initial letter that had a transparent sound mapping. There were no pictures whose target words began with diagraphs, e.g., 'sh' or 'ch', as a consequence.

naming paradigm was employed, to re-assess the treated and control items. The control task was also administered. Each post-therapy assessment session lasted roughly one to one-and-a-half hours, with breaks in between tasks. All picture naming trials (baseline, therapy, and post-therapy) were implemented in PsychoPy 2.0 (Pierce et al., 2019).

Standardised assessments were also administered during baseline sessions. These included (also see Appendix 5.1 for descriptions of each assessment):

(1) Language: As the primary aim was to explore the processes involved during orthographic cue-based anomia therapies, the selected assessments focused on verifying the participants' ability to interpret the written input and to produce a spoken output (also see Table 6.1, p. 168, for the complete list of language domains and standardised tests). Similar to the facilitation study (Chapters 3 and 4), the assessments were mainly subtests from the CAT (Swinburn et al., 2004) and PALPA (Kay et al., 1992) language batteries. The batteries were chosen as their test construction was grounded in an established psycholinguistic model (Kay et al., 1996). The Pyramid and Palm Trees Test (picture version, Howard & Patterson, 1992) was also included to replace similar tests in CAT or PALPA, as it contained a larger set of stimuli, and was thus potentially more able to detect subtle impairments related to conceptual semantics. Naming was assessed using the Boston Naming Test (Kaplan, Goodglass, and Weintraub, 2001). Finally, the Western Aphasia Battery Revised (WAB-R, Kertesz, 2007) was administered, as its Aphasia-Quotient provided a global score for the participants' language performance.

(2) Cognitive abilities: Scores from these standardised cognitive assessments (see Appendix 5.1 for brief test descriptions or Table 5.7 below for the list of assessments) could help us understand which cognitive areas were associated with any eventual therapy gains. These assessments were chosen for: (1) their common use within clinics and prior research, (2) their aphasia (especially, anomia)-friendliness, including minimal language demands particularly on the need for spoken/written output. They cover the key cognitive domains of attention (e.g., subtests in the Test of Everyday Attention (TEA); Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), memory (e.g., digit recall in Wechsler Memory Scale (WMS); Wechsler & Stone, 1972) and executive function (Brixton Spatial Anticipation Test; Burgess & Shallice, 1997), as well as general nonverbal reasoning (Raven's Coloured Matrices; Raven, 1976). In the case of digit recall and counting of elevator tones, where spoken responses of numbers were originally required, a number board was provided for the participant to point. Participants were assessed to make sure they could correctly identify the numbers beforehand (all participants scored 100% on spoken number-to-written number matching at baseline). They were also given the option to point and/or verbally report the

numbers during testing, according to their preferences. For each trial, the number board was only produced after the stimuli were presented (e.g., after the experimenter recited the numbers).

Finally, questionnaires (two parallel versions designed for participants and their caregivers respectively; see Appendices 5.2a and 5.2b for items) were administered during baseline session 3 as well as on each post-therapy assessment. The purpose of the bespoke questionnaires was to collect the participants'/caregivers' therapy-related feedback on possible qualitative changes in communication and other perceptions, which were not always detected by objective measures (e.g., naming accuracy). Ratings gathered will serve as supporting analyses (Section 5.3.1.2 below). Each questionnaire comprised of:

- (a) Two items on communication – one was on the rating of overall communication and the other, on spoken word retrieval. These aligned with the therapies' word-finding objective.
- (b) Three post-therapy items for the participants to rate their experience of therapies and indicate their preferred therapy; one post-therapy item for the caregivers about the ease of online setup.

5.2.3.2 Therapy sessions

Two cue therapies were compared in this study. Based on the A-B crossover design, some of the participants (three out of five participants, P3 to P5) received the 'Initial Letter: Orthographic cue' therapy first, before switching over to the 'Whole word: Orthographic cue' therapy. The reverse sequence was received by the remaining two participants (P1 and P2). The sequence each participant underwent was decided autonomously using Research Randomiser (Urbaniak & Plous, 2013), prior to the start of recruitment.

Like the facilitation study, bespoke target words were chosen for each participant. Efforts were made to select words participants could produce at least once during baseline naming, but this was not always possible if baseline performance was very poor. Importantly, five sets of target words were created (at least 24 items for each participant per condition, more details below) and these were randomly assigned into five conditions:

- Condition (1): Initial Letter Cue: Orthography only
- Condition (2): Repeated Exposure: Initial Letter Cue (i.e., Untreated items encountered during 'Initial Letter' orthographic cue therapy)
- Condition (3): Whole Word Cue: Orthography only

- Condition (4): Repeated Exposure: Whole Word Cue (i.e., Untreated items encountered during ‘Whole Word’ orthography cue therapy), and
- Condition (5): Unseen Control (i.e., Items only encountered during assessments but not during therapy)

Conditions (1) and (3) represented the orthographic cue therapies. The remaining conditions were control conditions (see p. 74 on why these conditions were necessary). All five sets of words were matched on: (a) the individual’s baseline naming performance, (b) subtitle contextual diversity (to index lexical frequency; Adelman et al., 2006), (c) number of letters, (d) number of phonemes, and (e) phonological consistency (Balota et al., 2007). One-way within ANOVA was performed on each of these five variables; results confirmed that matching was adequate [$p > .05$ for each variable].

In each therapy session, the participant received both the cue therapy condition and its corresponding repeated exposure condition. For example, when P2 underwent ‘Initial Letter’ orthographic cueing, conditions (1) and (2) would be administered. Each condition was delivered as a block to mitigate confusion. The order in which the two conditions were given, were randomised again using Research Randomiser and balanced across the 10 sessions allocated for each therapy.

Similar to the facilitation study, an additional subset of words (about 25% of the cued items) that the participant successfully named across all three baseline sessions, were embedded into each condition to encourage participants. These “items for encouragement” were excluded from further analyses because they were external to the matching between sets of treated items. The final item composition was: 33 items (26 were treated items) per condition for P1, 30 items (24 treated items) per condition for P2, 31 items (25 treated items) per condition for P3, 35 items (28 treated items) per condition for P4 and P5 respectively.

All therapy sessions were delivered on skype. For each therapy, 10 daily sessions (maximum 30-minutes each) were held across two weeks (weekends excluded). In each session, the researcher screen-shared the naming trials implemented on PsychoPy.

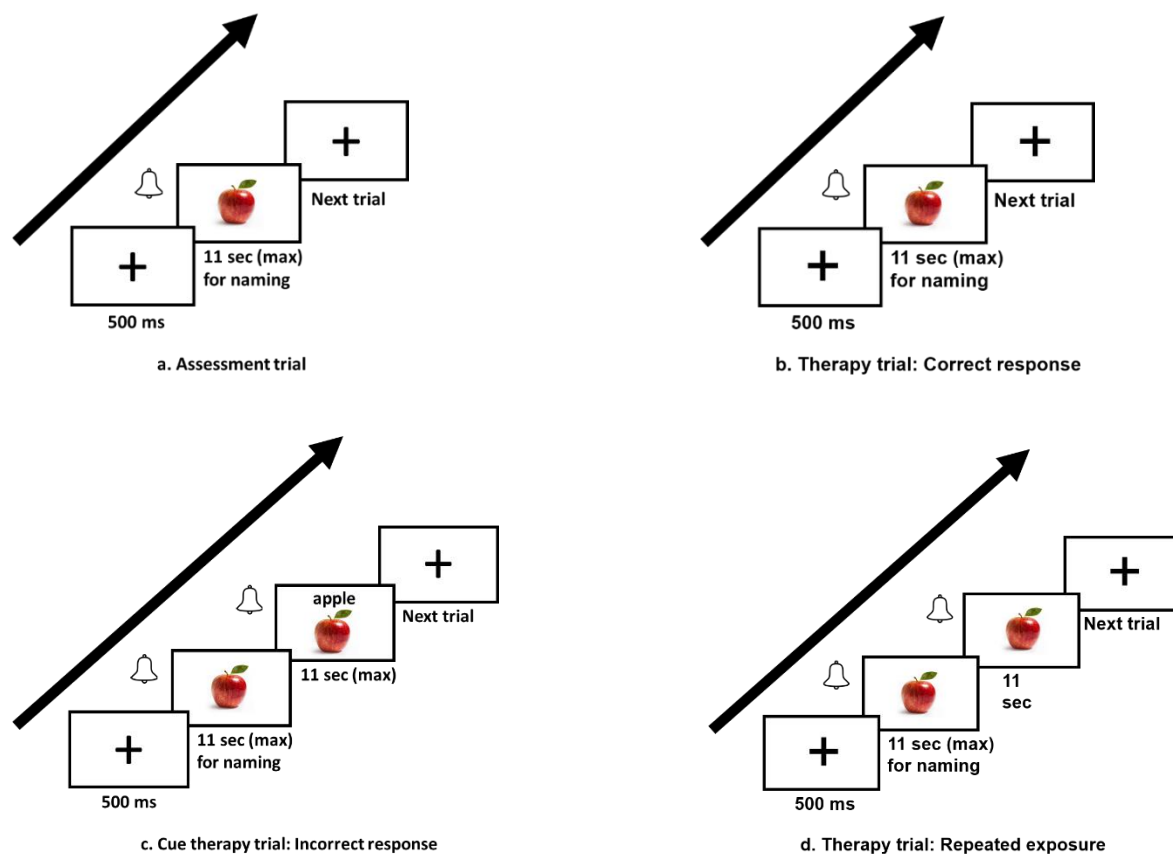
5.2.3.3 Picture-naming trials during therapy

Figures 5.2b-d show the breakdown of the picture-naming trials. The sequence was similar to that of the facilitation study, except this time, only written forms (no phonological prompts) were provided as cues. Every trial begun with a 500-ms fixation cross, which directed the participant to the screen. The picture was then displayed, with an alerting sound simultaneously played. The picture remained for a maximum of 11 seconds, during

which the participant would try name it. When correctly named, the picture cut to the next trial. When no correct responses were produced within 11 seconds, an orthographic cue (either the first letter or the entire word string) was presented above the picture, which stayed online for an additional 11 seconds before proceeding to the next trial. Participants were free to process the written cues in any manner they preferred (e.g., sound them aloud or read silently). Similar to the facilitation study (Chapters 3 and 4), there is no explicit feedback on accuracy for each trial; feedback is implicit because an incorrect response will result in extra time and a cue to assist with the naming (Figure 5.2c).

Each picture was named once per session. The order of the items was randomised every session. Finally, there were five practice trials at the start of every session. Stimuli used in practice trials were neither treatment nor control items.

Figure 5.2. Format of assessment and therapy trials.



5.2.3.4 Extraction of main naming outcome measures and fidelity check

All spoken picture-naming responses during assessments were audio-recorded using the Sony ICDPX370 Mono Digital Voice Recorder.

The three main naming outcome measures previously analysed in Chapters 3 and 4 were also extracted from the audio recordings here, i.e., initial naming accuracy, final naming accuracy, and response times for correct responses (please refer to Section 3.2.4.5 on p. 76 for the operationalisations).

A certified speech and language therapist fluent in English (this individual was different from the rater who assisted with the facilitation study) external to the research team, independently verified 20% of the data based on the three main outcome measures. These data were randomised for the rater, who was therefore blind to the exact condition assigned to each item, as well as the point in the study which the data arose. The previous steps taken to settle differences and to compute coding agreements were also replicated (see p. 77 for the procedure). Overall, interrater agreement was high across all three primary outcome variables at each time point, with at least 96% percentage agreement on each measure for each assessment session [Average percentage agreement across sessions: Initial naming accuracies = 99.09%, final naming accuracies = 99.60%, RT = 97.78%] and with Krippendorff's α per measure per session exceeding .80 [Average α -values across sessions: Initial naming accuracies = .991; Final naming accuracies = .996; RT = .978](Krippendorff, 2004; Freelon, 2010, 2013).

5.2.3.5 Rating outcomes for participants and caregivers' therapy-related views

Questionnaire responses were rated on a six-point aphasia-friendly Likert scale (a six-point single-item rating scale was also used to gauge aphasia severity in Goodglass, Kaplan, and Barresi, 2001; also see Chomeya, 2010). Care was taken to ensure the questionnaires (Appendices 5.2a and 5.2b) were completed by participants and caregivers independently, to avoid mutual influence. All ratings were first reverse-coded before analysis, such that a higher numerical rating would indicate a more favourable outcome.

5.2.3.6 Ethics clearance and impact of Covid-19

Ethics was submitted and departmental clearance was received (Project ID: LCD-2019-05), before recruitment commenced. Data collection was ongoing when the Covid-19 pandemic struck. Ethics, adjusted for Covid-19, was resubmitted and cleared. The initial plan was to administer all assessment sessions (baseline and post-therapy) in-person face-to-face. When Covid-19 lockdown restrictions were imposed, all sessions including assessments had to be transferred onto the virtual platform. This change mainly affected

the last three participants (P3-P5), because the first two were seen before the pandemic broke out in UK. On reflection, the impact on the study was not huge and the transition, generally smooth. Although there were unavoidable concerns over the uncertainties and inconveniences created by Covid-19, specific to the implementation of this study, there were no major IT issues and sessions proceeded relatively uninterrupted. P4 required on-site support related to equipment setup and this was resolved by the experimenter on a home visit, just before lockdown was enforced. Perhaps because pre-Covid, the original therapy was conceived to be an online experience conducted over skype, the study might have primarily attracted individuals coming from households that were IT/internet-ready, and/or whose live-in caregivers were to some extent, IT-competent. Shifting everything online therefore incurred no severe disruption.

5.2.4 Group-level analyses

Primary analyses (research questions 1 and 2) were carried out across participants and across the three time points, for each of the three main outcome measures (initial and final naming accuracy, response time). At each time point and for each condition, mean proportional outcomes, i.e., average naming accuracies and average RT across items, were computed at the participant-level and these values were entered as the dependent variables.

The three time points²⁸ were:

- (a) the baseline session closest to the start of each therapy (referred to as the ‘final baseline’),
- (b) one-week post-therapy, and
- (c) four weeks post-therapy.

These time points were used because they shared same temporal intervals (between pre- to post-therapy) across the conditions, and were therefore most comparable, notwithstanding the imbalance in number of data points between therapies created by the A-B crossover design (see Figure 5.1)²⁹. In view of the small sample, non-parametric tests

²⁸ If the ‘Initial Letter: Orthography only’ cue therapy was administered first before the ‘Whole Word: Orthography only’ cue therapy, the final baseline or baseline closest to/just prior to the initial letter therapy was ‘baseline session 3’. The baseline closest to the whole word therapy was ‘post-therapy session 1’ (see Figure 5.1). For initial letter therapy, one-week post-therapy was ‘post-therapy session 1’ and four weeks post-therapy, being ‘post-therapy session 2’. For whole word therapy, one-week post-therapy was ‘post-therapy session 2’ and four weeks post-therapy, being ‘post-therapy session 3’.

²⁹ Data from seven weeks after the first therapy was not included in the group analysis, as there was no comparable data from seven weeks after the second intervention.

were used as per Chapter 3. Friedman ANOVA evaluated the overall effect across these three time points. Post hoc testing on significant findings was conducted using Wilcoxon signed-ranks tests.

Supporting analyses based on questionnaire ratings were analysed using planned pairwise comparisons with Wilcoxon signed ranks tests. Across respondents (participants and caregivers analysed separately), each post-therapy rating was evaluated against its baseline rating.

For the secondary analysis (secondary research question 3), an additional variable ‘therapy proportional gain’ (referred to as ‘therapy gain’ for short) was computed for each of the three main outcome measures at the one and four-weeks post-therapy time points. The computation followed the formula for ‘facilitation proportional gain’ described in Chapter 3 (see Section 3.2.4.6.2 on p. 79 for details, also refer to p. 294 in Lambon Ralph et al., 2010, for a similar operationalisation), i.e., for initial and final naming accuracies: Subtract the final baseline score from the post-therapy scores respectively, before dividing by the number of treated items still not successfully named by the final baseline; and for RT: Subtract the post-therapy RT from the baseline RT, before dividing by the baseline RT. The final baseline referred to the baseline score immediately prior to each intervention. These values indicated the amount of progress the participant made and retained from the therapy, proportional to the potential maximum given the treatment set size/RT scope.

Kendall’s rank order correlation was then run across participants; their ‘therapy gains’ for each condition at each time point were correlated with the proportional accuracy scores achieved for various cognitive assessments at baseline. Similar to Chapter 3 (p. 81), proportional accuracy was operationalised as the total score obtained at baseline divided by the maximum possible score for that assessment (i.e., accuracy scores are defined in terms of correct responses, not the number of errors). Kendall’s rank order correlation was appropriate, given the small sample size (Siegel & Castellan Jr., 1988).

5.3 Results

The results are reported in the order of the research questions. The focus of this chapter will be results from group-level analyses. Individual analyses are reported in the next chapter.

5.3.1 Primary research questions (1) and (2): At the participant group level, which unit of the orthographic cue will result in significant improvement; also, how durable are the therapy effects of each orthographic cue?

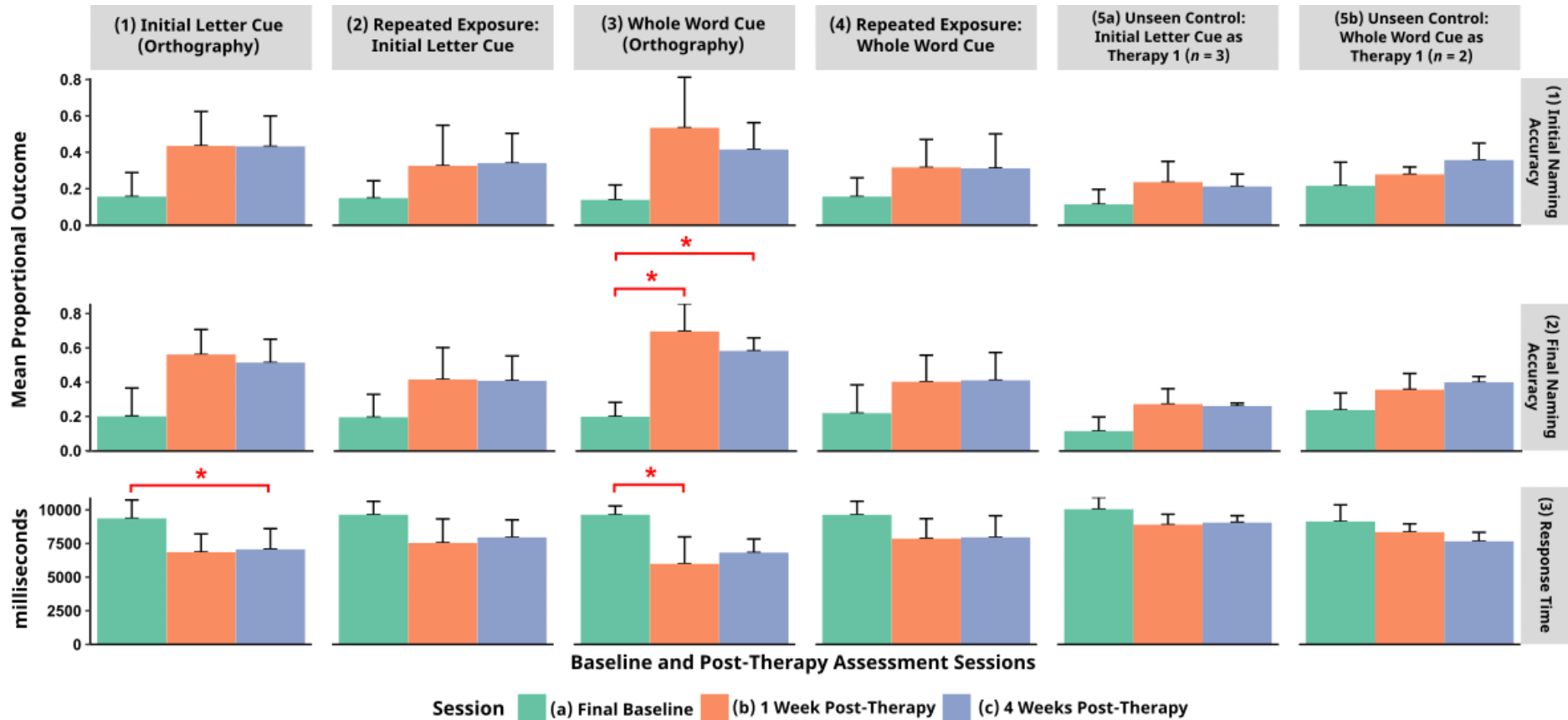
5.3.1.1 Primary analyses on the main naming outcome measures

The descriptive statistics are shown in Figure 5.3. The complete results are published in Tables 5.3 to 5.5. Performances on unseen control items were analysed separately based on the first cue therapy (Therapy 1) that the participant completed. This was again a corollary of the A-B crossover design – unseen control items should capture the transfer effects from the therapy that was most recently administered; thus, it would not be fair to equate performance on the unseen control condition tested directly after ‘Initial Letter: Orthography’ cueing ($n = 3$) to the analogous control condition after ‘Whole Word: Orthography’ ($n = 2$) cueing, since their transfer effects could differ. Lastly, all findings are consolidated in Table 5.6 to ease interpretation, with all significant findings rendered yellow.

Except for the unseen control condition, group-level Friedman analyses revealed significant improvements in all four remaining conditions across the three time points, on all three outcome measures. Post-hoc comparisons using Wilcoxon signed ranks tests further showed that, in these four conditions, the scores respectively achieved one and four weeks post-therapy were significantly higher than their corresponding baseline scores. To account for multiple testing during post hoc comparisons, LSD procedure was applied (see p. 78 for details).

Given that repeated exposure also resulted in significant improvements (Tables 5.3-5.6), additional Wilcoxon signed ranks tests were run to verify that significant improvements identified for the two cue therapies were not simply a function of repeated encounters with the same pictures (see Nickels, 2002a, Creet et al., 2019). When compared against significant results from their corresponding repeated exposure conditions, effects from initial letter cues remained significant only for RT four weeks later. In comparison, effects from the written whole word cues remained significant in three outcomes: (a) Final naming accuracy one week later, (b) final naming accuracy four weeks later (marginal), and (c) RT one week later. Just to confirm that the above evaluations between cue conditions and repeated exposure conditions were not biased by discrepancies at baselines, baseline checks were conducted. This was achieved by comparing baseline performances measured on each outcome, just prior to the therapy, using Wilcoxon Sign Ranks tests. Results from the baseline checks were not significant (Tables 5.3 - 5.5).

Figure 5.3. Descriptive statistics of group-level results ($n = 5$) for each of the conditions and the three main naming outcome measures at the final baseline and post-therapy sessions (one week and four weeks post-therapy). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated.



+ .05 $< p \leq .07$, * .01 $< p \leq .05$, ** .001 $< p \leq .01$, *** $p \leq .001$

Table 5.3. Results of the group analysis by-participants based on the mean proportional accuracy of their initial naming responses.

Condition	Friedman ANOVA	1-Week Post-Therapy vs. Final Baseline	4-Weeks Post-Therapy vs. Final Baseline
Initial Letter Cue: Orthography	* $\chi^2(2, N = 5) = 7.60$, $p = .022$, Kendall's $W = .76$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Initial Letter Cue: Orthography)	* $\chi^2(2, N = 5) = 6.63$, $p = .036$, Kendall's $W = .66$	† $z = -1.83, p = .068$, $r = -0.58$	* $z = -2.02, p = .043$, $r = -0.64$
Post hoc (if above results are sig): Initial Letter against Rep Exp.	BL check (*): <i>n.s.</i> $z = -0.45, p = .66$ $r = -0.14$	<i>n.s.</i> $z = -1.48, p = .14$, $r = -0.47$	<i>n.s.</i> $z = -1.75, p = .0796$, $r = -0.55$
Unseen Control: Untreated, Unexposed (Initial Letter Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 3) = 4.91$, $p = .086$, Kendall's $W = .82$		
Whole Word Cue: Orthography	* $\chi^2(2, N = 5) = 6.42$, $p = .040$, Kendall's $W = .64$	† $z = -1.83, p = .068$, $r = -0.58$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Whole Word Cue: Orthography)	* $\chi^2(2, N = 5) = 6.63$, $p = .036$, Kendall's $W = .66$	* $z = -2.02, p = .043$, $r = -0.64$	† $z = -1.83, p = .068$, $r = -0.58$
Post hoc (if above results are sig): Whole Word Cue against Rep Exp.	BL check (*): <i>n.s.</i> $z = -0.37, p = .72$ $r = 0.12$	<i>n.s.</i> $z = -1.46, p = .14$, $r = -0.46$	<i>n.s.</i> $z = -1.75, p = .0796$, $r = -0.55$
Unseen Control: Untreated, Unexposed (Whole Word Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 2) = 3.71$, $p = .16$, Kendall's $W = .93$		

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

(*) Baseline check: Performance at baseline immediately prior to therapy (e.g., For therapy 2 which commenced on Week 8, its 'final baseline' would be the baseline assessment collected at Week 7) versus corresponding performance at baseline (also Week 7) for repeated exposure condition. Tested using McNemar Sign Ranks test.

Table 5.4. Results of the group analysis by-participants based on the mean proportional accuracy of their final naming responses.

Condition	Friedman ANOVA	1-Week Post-Therapy vs. Final Baseline	4-Weeks Post-Therapy vs. Final Baseline
Initial Letter Cue: Orthography	* $\chi^2(2, N = 5) = 8.32$, $p = .016$, Kendall's $W = .83$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Initial Letter Cue: Orthography)	* $\chi^2(2, N = 5) = 7.60$, $p = .022$, Kendall's $W = .76$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Post hoc (if above results are sig): Initial Letter against Rep Exp.	BL check (*): <i>n.s.</i> $z = -0.45, p = .65$ $r = -0.14$	<i>n.s.</i> $z = -1.21, p = .22$, $r = -0.38$	<i>n.s.</i> $z = -1.604, p = .11$, $r = -0.51$
Unseen Control: Untreated, Unexposed (Initial Letter Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 3) = 4.91$, $p = .086$, Kendall's $W = .82$		
Whole Word Cue: Orthography	* $\chi^2(2, N = 5) = 8.32$, $p = .016$, Kendall's $W = .83$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Whole Word Cue: Orthography)	* $\chi^2(2, N = 5) = 6.63$, $p = .036$, Kendall's $W = .66$	† $z = -1.83, p = .068$, $r = -0.58$	* $z = -2.02, p = .043$, $r = -0.64$
Post hoc (if above results are sig): Whole Word Cue against Rep Exp.	BL check (*): <i>n.s.</i> $z = -0.37, p = .72$ $r = 0.12$	* $z = -2.02, p = .043$, $r = -0.64$	† $z = -1.84, p = .066$, $r = -0.58$
Unseen Control: Untreated, Unexposed (Whole Word Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 2) = 3.71$, $p = .16$, Kendall's $W = 0.93$		

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

(*) Baseline check: Performance at baseline immediately prior to therapy (e.g., For therapy 2 which commenced on Week 8, its 'final baseline' would be the baseline assessment collected at Week 7) versus corresponding performance at baseline (also Week 7) for repeated exposure condition. Tested using McNemar Sign Ranks test.

Table 5.5. Results of the group analysis by-participants based on the average naming response times.

Condition	Friedman ANOVA	1-Week Post-Therapy vs. Final Baseline	4-Weeks Post-Therapy vs. Final Baseline
Initial Letter Cue: Orthography	* $\chi^2(2, N = 5) = 7.60$, $p = .022$, Kendall's $W = .76$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Initial Letter Cue: Orthography)	* $\chi^2(2, N = 5) = 8.40$, $p = .015$, Kendall's $W = .84$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Post hoc (if above results are sig): Initial Letter against Rep Exp.	BL check (*): <i>n.s.</i> $z = -1.21, p = .22$ $r = -0.38$	<i>n.s.</i> $z = -0.67, p = .50$, $r = -0.21$	* $z = -2.02, p = .043$, $r = -0.64$
Unseen Control: Untreated, Unexposed (Initial Letter Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 3) = 4.67$, $p = .097$, Kendall's $W = .78$		
Whole Word Cue: Orthography	* $\chi^2(2, N = 5) = 7.60$, $p = .022$, Kendall's $W = .76$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Repeated Exposure (Whole Word Cue: Orthography)	* $\chi^2(2, N = 5) = 7.60$, $p = .022$, Kendall's $W = .76$	* $z = -2.02, p = .043$, $r = -0.64$	* $z = -2.02, p = .043$, $r = -0.64$
Post hoc (if above results are sig): Whole Word Cue against Rep Exp.	BL check (*): <i>n.s.</i> $z = -0.13, p = .89$ $r = 0.043$	* $z = -2.02, p = .043$, $r = -0.64$	<i>n.s.</i> $z = -1.48, p = .14$, $r = -0.50$
Unseen Control: Untreated, Unexposed (Whole Word Cue as Therapy 1)	<i>n.s.</i> $\chi^2(2, N = 2) = 4.00$, $p = .14$, Kendall's $W = 1.00$		

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

(*) Baseline check: Performance at baseline immediately prior to therapy (e.g., For therapy 2 which commenced on Week 8, its 'final baseline' would be the baseline assessment collected at Week 7) versus corresponding performance at baseline (also Week 7) for repeated exposure condition. Tested using McNemar Sign Ranks test.

Table 5.6. Summary of group-level findings on the effectiveness of each condition on the three main naming outcome measures.

Condition		<u>Initial Letter Cue (Ortho)</u>			<u>Repeated Exposure: Initial Letter Cue (Ortho)</u>			<u>Whole Word Cue (Ortho)</u>			<u>Repeated Exposure: Whole Word Cue (Ortho)</u>			<u>Unseen Control: Untreated Unexposed</u>		
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT
Group (By Pt)		*	*	*	*	*	*	*	*	*	*	*	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against final baseline)	x1 Week. Post- vs. BL	*	*	*	†	*	*	†	*	*	*	†	*	—	—	—
	x4 Weeks Post- vs. BL	*	*	*	*	*	*	*	*	*	†	*	*	—	—	—
Post-hoc (against rep. exposure)	Final Baselines	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>				<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>				—	—	—
	x1 Week Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>				<i>n.s.</i>	*	*						
	x4 Weeks Post	<i>n.s.</i>	<i>n.s.</i>	*				<i>n.s.</i>	†	<i>n.s.</i>						
		(<i>p</i> = .0796)						(<i>p</i> = .0796)								

† .05 < *p* < .07, * .01 < *p* < .05, ** .001 < *p* < .01, *** *p* < .001

Finally, for clinicians interested in quantifying improvement using raw figures, all group-level changes in terms of mean raw values can be located in Appendix 5.3. The outcomes that remained significant after controlling for effects from repeated exposure are highlighted as follow: On average, after receiving 10 sessions of initial letter orthographic cueing, participants were faster by 2287.88 ms (change in raw values reflecting a proportional change of 0.24) with a 95% confidence interval of [824.22 ms, 3751.54 ms]. In contrast, after receiving 10 sessions of whole word orthographic cueing, participants on average were able to name an additional 13.00 words (proportional change of 0.61) with a 95% confidence interval of [7.51, 18.49] one week later, and 10.00 more words (proportional change of 0.48) with a 95% confidence interval of [8.11, 11.89] four weeks later. These were based on comparisons against performance on the most immediate baseline prior to treatment, using final naming accuracy (Table 5.6; Appendix 5.3). On average, after receiving the whole word orthographic cue therapy, participants were also faster by 3647.54 ms (proportional change of 0.38) with a 95% confidence interval of [1616.14 ms, 5678.95 ms] at one-week post-therapy.

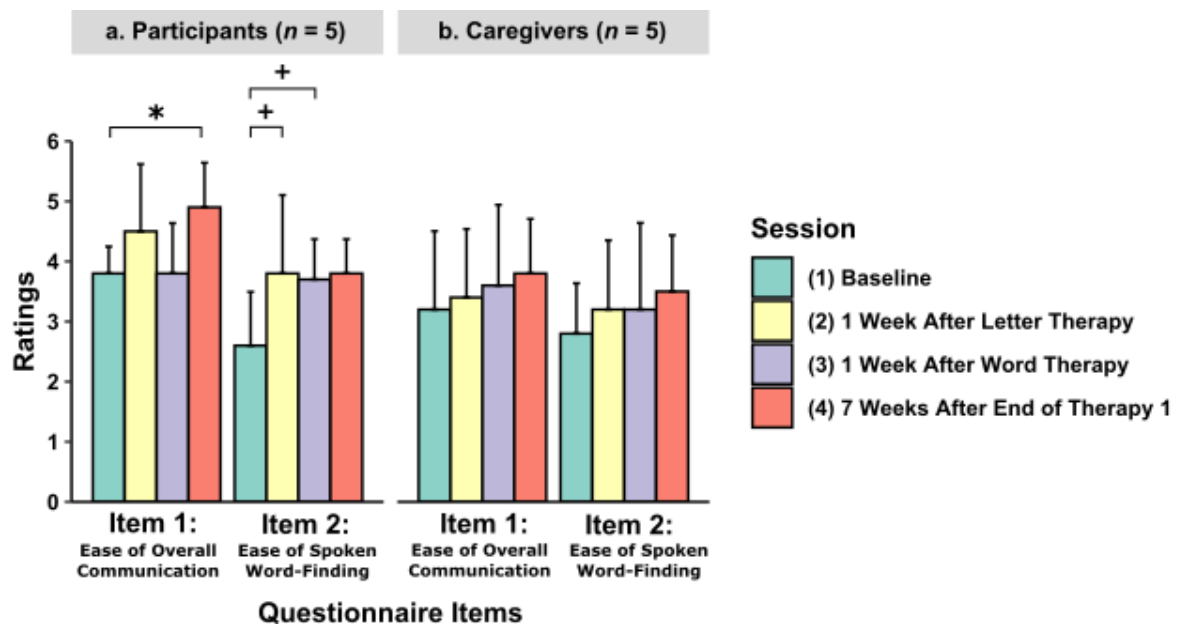
5.3.1.2 Primary supporting analyses on rating outcome measures (Questionnaire responses)

Mean ratings were computed across respondents (responses from participants and caregivers were calculated separately) for the first two items. Descriptive statistics for their responses are depicted in the bar graphs below (Figure 5.4). The results for each item were analysed using planned pairwise comparisons with Wilcoxon signed ranks tests, such that each rating at the three post-therapy time point was evaluated against its baseline rating. This generated three comparisons for each item.

Qualitatively, across the four time points, a general upward trend of post-therapy improvement was discerned based on ratings by both participants and their caregivers. Pairwise comparisons revealed that for the participants, compared to the baseline, there was a significant improvement on their self-reported overall communication ability on the seventh week after the start of the first therapy [$z = -2.06$, $p = .039$, $r = -0.65$]. Specific to spoken word-finding, marginal improvements were also reported by the participants one week post-‘Initial Letter: Orthography’ cue therapy [$z = -1.89$, $p = .059$, $r = -0.60$] and one week post-‘Whole word: Orthography’ cue therapy [$z = -1.84$, $p = .066$, $r = -0.58$], compared to the baseline. The rest of the comparisons were not significant.

For the caregivers, pairwise comparisons did not detect significant pre- versus post-therapy improvement, be it on their perception of the participants' overall communication or just their spoken language [all three p s for each item > .10].

Figure 5.4. Descriptive statistics of ratings (scale of '1' to '6') on items 1 and 2 from participants and caregivers.



[†] .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

In terms of therapy experience, out of a maximum of '6' ('6' being 'excellent' after recoding), the average participant rating for 'Initial Letter: Orthography' cue therapy was '5.4' ($SD = .89$), for 'Whole Word: Orthography' cue therapy was '5.1' ($SD = .74$). However, when asked to choose between the two therapies on the final post-therapy assessment session, four participants preferred the whole word written cue therapy, and one (P4) could not decide because he liked both initial letter and whole word written cue therapies equally. When verbally asked which of the two therapies was less frustrating, all five participants, indicated by way of pointing, the whole word written cue therapy.

As for the ease of using a virtual platform like Skype to receive anomia therapy ('6' being 'very easy' after data recoding), participants gave an average rating of '5' ($SD = .71$), '4.8' ($SD = 1.30$), and '5.2' ($SD = 1.30$) respectively during the three post-therapy sessions.

Out of the five caregivers, three assisted with the daily initial Skype log-in/set up. These were the caregivers of P1, P2, and P5; they rated the ease of setting up the online sessions as '5.83' ($SD = .29$) and '5.33' ($SD = .58$) at the end of the participants' initial letter and whole word written cue therapies respectively.

5.3.2 Secondary research question (3): At the participant group level, what is the relationship between baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?

Results of Kendall's rank order correlation between baseline cognitive scores and proportional gains from each orthographic therapy cue at each time point, are presented in Table 5.7.

At one-week post therapy, there were significant positive and strong correlations between therapy gains (Initial naming accuracy and RT) for whole word written cues and 'Raven's Coloured Progressive Matrices' [in both cases: $\tau_b = .95$, $p = .023$]. The correlation between 'Elevator Counting with Distraction: TEA Subtest 3') and initial naming therapy gains following initial letter written cues was marginally significant [$\tau_b = .84$, $p = .052$]. The rest of the results were not significant, although the correlation between 'Raven's Coloured Progressive Matrices' and final accuracy outcome following whole word written cues approached marginal significance and the strength was strong [$\tau_b = .74$, $p = .077$].

At four-weeks post therapy, three significant correlations were present. For initial letter written cues, there was (1) a strong positive relationship between RT gains and 'Raven's Coloured Progressive Matrices' [$\tau_b = .95$, $p = .023$]. As for whole word cues, there were strong positive relationships between 'Raven's Coloured Progressive Matrices' with (2) final naming accuracy and (3) RT gains respectively [in both cases: $\tau_b = .95$, $p = .023$]. There was also a marginally significant negative correlation between the final naming accuracy therapy gains following the initial letter cue, and 'Elevator counting: TEA Subtest 2' [$\tau_b = -.84$, $p = .052$]. In addition, the correlation between 'Raven's Coloured Progressive Matrices' and initial accuracy outcome following initial written cues also approached marginal significance [$\tau_b = .74$, $p = .077$].

5.3.3 Control task: Nonword repetition

On average, the participants scored 13.40 ($SD = 6.39$) at baseline, 16.00 ($SD = 5.10$), 15.80 ($SD = 4.15$), and 15.60 ($SD = 5.59$) at post-therapy assessment sessions 1, 2,

Table 5.7. Kendall rank-order correlations between performance on baseline cognitive assessments and the proportional naming gains yielded across different post-therapy time points (i.e., Initial Naming (IN): Therapy Gain at 1 week and 4 weeks (1 Wk, 4 Wks), Final Naming (FN): Therapy Gain at 1 week and 4 weeks (1 Wk, 4 Wks), and Response Time (RT): Therapy Gain at 1 week and 4 weeks (1 Wk, 4 Wks)) for each cue condition.

	Initial Letter Cue: Orthography Only						Whole Word Cue: Orthography Only					
	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk	IN: 4 Wks	FN: 4 Wks	RT: 4 Wks	IN: 1 Wk	FN: 1 Wk	RT: 1 Wk	IN: 4 Wks	FN: 4 Wks	RT: 4 Wks
Attention												
Elevator Counting (TEA Subtest 2)	-.12	-.60	-.60	-.60	-.84[†] (<i>p</i> = .052)	-.36	-.36	-.60	-.36	.12	-.12	-.12
Elevator Counting with Distraction (TEA Subtest 3)	.84 (<i>p</i> = .052)	.36	.36	.36	.12	.12	.12	.36	.12	.60	.36	.36
Memory: Short Term Memory												
Recognition Memory (CAT Subtest 4)	0	0	0	.52	.52	.26	.26	.52	.26	0	0	0
Corsi Block Tapping Test	0	.45	.22	0	.22	-.22	-.22	0	-.22	-.45	-.45	-.45
Digits Forward (Wechsler Memory Scale II)	-.26	-.77	-.77	-.26	-.52	-.26	-.26	-.26	-.26	.26	-.26	-.26
Memory: Working Memory												
Digits Backward (Wechsler Memory Scale II)	.60	.12	.12	-.12	-.36	-.12	-.12	-.12	-.12	.60	.12	.12
Executive Function												
Brixton Spatial Anticipation Test	-.11	-.32	-.11	.53	.32	.74 (<i>p</i> = .077)	.74 (<i>p</i> = .077)	.53	.74 (<i>p</i> = .077)	-.11	.53	.53
General Nonverbal Reasoning												
Raven's Coloured Progressive Matrices	.32	.11	.32	.74 (<i>p</i> = .077)	.53	.95* (<i>p</i> = .023)	.95* (<i>p</i> = .023)	.74 (<i>p</i> = .077)	.95* (<i>p</i> = .023)	.32	.95* (<i>p</i> = .023)	.95* (<i>p</i> = .023)

† .05 < *p* < .07, * .01 < *p* < .05, ** .001 < *p* < .01, *** *p* < .001

and 3 respectively. Based on Friedman ANOVA, there was no significant change in their nonword repetition accuracies across time [$\chi^2(3, n = 5) = 6.23, p = .101$, Kendell's $W = .42$].

5.4 Discussion

In this chapter, three research questions were investigated using group-level data. Attempts were made to adjudicate between the therapeutic efficacy of two units of orthographic cues, initial letter versus whole word, by considering their therapy outcomes (Research question 1) and the length of their effects (Research question 2). Primary analyses were run on the three primary main outcome measures (initial and final naming accuracies, RT). These were supplemented by supporting analyses of the participant/caregiver ratings for their views on the spoken word-finding therapies received and impact on daily communication. Associations between cognitive abilities and the therapy benefits from each orthographic cue at various post-therapy time points were also explored (Research question 3).

Results from the primary analyses showed that both orthographic initial letter and whole word cues resulted in significant naming improvements at one- and four-weeks post-therapy respectively on all three main measures (Tables 5.3 to 5.6). However, when compared against effects from repeated exposure, effects from initial letter orthographic cues remained significant on average RT (four weeks post-therapy), but on the two accuracy measures. In comparison, significant effects from whole word orthographic cues persisted both on final naming accuracy (one- and four weeks post-therapy) and average RT (one-week post-therapy). Consistent with literature, these effects were item-specific (e.g., Miceli et al., 1996; Basso et al., 2001), with no significant generalisation found on untreated items and no general improvement detected (there was no significant change in performance for the control task). This set of findings is important as firstly, these are novel (see Table 5.1 for a summary on previous studies related to orthography-based anomia therapies): They (1) provided new evidence on the efficacy of both initial letter and whole word orthographic cues and in particular, (2) demonstrated that ten applications of a whole word orthographic cue over a course of two weeks could result in improved naming accuracy sustained for at least a week and potentially, a month. Quantitatively, the improvements from whole word orthographic cues were also substantial: there was an average naming gain of 13.00 additional words (proportional change of 0.61) compared to pre-therapy when assessed one week later, and 10.00 more words (proportional change of 0.48), four weeks later. In

addition, the participants also significantly enhanced their average naming speed by 3647.54 ms (proportional change of 0.38), when they were assessed at one-week post-therapy.

Secondly, these findings were validated against the effects of repeated exposure, which constitute a threat to the internal validity of therapy studies. Repeated exposure are priming effects following repeated encounters with an item, which would naturally arise for any adult participants with developed language systems, even if the systems were compromised after a stroke (Creel et al., 2019). Presently, not all therapy studies accounted for these effects (e.g., COMPARE trials, Rose et al., 2021). It is important to note that comparisons against repeated exposure constitute a tough criterion; moreover, these analyses are further statistically disadvantaged by the small sample size (i.e., Type II *not* Type I error would be more probable given the conservatism).

Supporting analyses detected modest improvements on spoken word-finding, based on questionnaire ratings reported by the participants one-week after each written cue therapy ($p = .059$ and $.066$ respectively). On their final post-therapy assessment (i.e., on post-therapy assessment 3), participants rated their overall communication to have had improved significantly from baseline ($p = .039$). Although the changes in ratings from caregivers were not significant; qualitatively, a similar trend of improvement was elicited (Figure 5.4b).

Importantly, average ratings by participants on their online therapy experience were also high (> 5' out of a maximum of '6'). Both participants and caregivers rated the use of an online platform, like Skype, to be relatively easy (participants' and caregivers' average ratings were respectively around '5' out of a maximum of '6'). Together, these high approval ratings suggested that our older adults found the daily online orthographic cue therapies a feasible routine for them to implement, with participants themselves perceiving beneficial communication outcomes afterwards (also see General Discussion, Section 8.8, p. 234, on the implications of online language therapy, especially during lockdown emergencies).

Finally, correlation results from the secondary analyses showed that naming therapy gains (naming accuracies and RT) on both initial letter and whole word orthographic cues were significantly related to nonverbal reasoning skills captured by 'Raven's Coloured Progressive Matrices' (Table 5.7) at various time points. These findings suggested that there may likely be shared skills between general cognitive reasoning and the capacity to benefit from an orthographic cue (initial letter or whole word). One commonality would be the interpretation of visual stimuli and use of the visual information extracted to complete a task (be it picture-naming or solving a puzzle). However, given the small sample size, this finding is at best exploratory. The general association between nonverbal reasoning and

beneficial therapy outcomes is in keeping with other larger-scale findings involving anomia cue therapies (e.g., Rose et al., 2021, who also reported performance on Raven's Progressive Matrices as a predictor of the change in naming accuracy for treated items for individuals with mild aphasia). Further research is required to clarify the nature of this relationship.

In conclusion, group findings from this therapy study supported the use of initial letter and whole word orthographic cues to treat spoken word-finding difficulties. Notably, across the participants, whole word orthographic cues delivered a slight advantage, improving *both* accuracy and RT outcomes. There is further supporting evidence for whole word orthographic cues from the participants' feedback – four out of the five participants preferred the whole word written cue over the initial letter prompt. In the next chapter, these outcomes are considered at the individual level.

Chapter 6

Comparing the use of initial letter against whole word as the unit of orthographic cue in spoken naming therapy – Individual participant analyses

6.1 Introduction

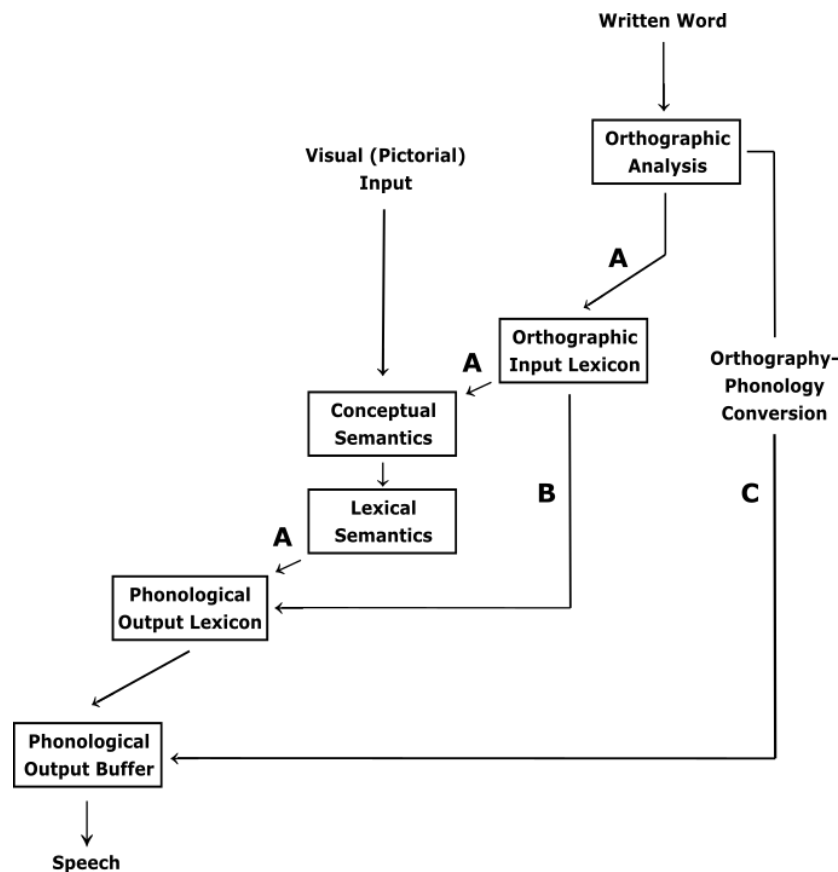
Taking into account individual differences during intervention is important to patient-centred care (Schork, 2015). Individual-level analyses enable investigation into how inter-individual differences affect treatment outcomes. In this chapter, our major goal is to go beyond the average responses obtained from group analysis in Chapter 5, by inspecting each participant's responses in turn (see Best, Sze, Edmundson, & Nickels, 2019, or p. 97 in Chapter 4 on a review on how group averages can mask individual differences).

One of the primary findings in Chapter 5 was the overall superiority of whole word written cues at the group level, with indications that their therapeutic effects could last for at least a week and potentially four weeks, after 10 sessions of cue therapy. On the other hand, effects from the initial letter written cues were not significant, when evaluated against effects of repeated exposure. It is therefore important to follow-up and identify possible individual differences within these outcomes, specifically which participant(s) benefitted from which type(s) of written cue(s) and for how long, as well as whether their language and cognitive profiles could help explain therapy outcomes.

To explore which language abilities were associated with a stronger capacity to benefit from a particular unit of orthographic cue, references will be made to the model illustrated in Figure 6.1 (see previous discussion on a fuller version model that incorporates auditory processing components on pp. 5-7 in Chapter 1 and p. 98 in Chapter 4; also refer to Colheart, 2012, for a summary of related models on reading aloud). Figure 1.2 is adapted from neuropsychological models, including the one underpinning PALPA (Kay et al., 1996). Figure 6.1 represents a more distilled version that focuses on elements related to orthographic processing. To recap, the orthographic processing pathways of interest are: Path A (the lexical-semantic route activated when there is a need to understand the meaning of the written word), path B (the direct lexical route that allows words one recognises be read read aloud, even if their meanings are not understood and if their spellings are irregular) and path C (the sublexical route that assumes a letter string has regular letter-sound mappings as it converts the graphemes into their corresponding phonemes for reading aloud, and is applicable for nonwords or novel words an individual has not previously encountered). The aim here is to explore and extract systematic patterns within participants' baseline data

according to these three processing pathways, and to determine whether these patterns could be systematically related to that individual's ability to benefit (or not benefit) from specific orthographic cues.

Figure 6.1. Orthographic input processing (visual word/nonword processing) pathways with spoken output.



In the same vein, systematic patterns according to performances on different cognitive tasks (e.g., attention, memory, executive function) were explored to determine which could be associated with the individuals' abilities to profit from particular written cues. A systematic pattern is defined as a positive relationship linking an individual's ability to successfully benefit from a particular cue to his/her specific strength(s) in terms of orthographic input processing (Figure 6.1) and/or aspects of cognition. Conversely, a lack of ability to significantly benefit from a particular cue would be tied to specific deficit(s) in either orthographic input processing or cognitive abilities or both.

The research questions are therefore similar to Chapter 5's, but with an emphasis on identifying the outcomes of specific participants and how their individual

strengths/weaknesses relate to their performances post-therapy. Specifically, the following were asked:

- **Primary research question (1): At the individual participant level, which unit of the orthographic cue (initial letter versus whole word) will result in significant improvement on the three main naming outcomes (initial/final accuracies, RT), following multiple applications of each cue?**
- **Primary research question (2): How durable are the therapy effects of each orthographic cue (initial letter versus whole word) for each participant?**
- **Secondary research question (3): What is the relationship between each participant's baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?** Investigation into this will be exploratory.

6.2 Method

6.2.1 Design, procedure and stimuli

This study compared two orthographic cue therapies (initial letter against whole word written cue) using an A-B crossover within-participant case-series design across six time-points (three baseline sessions (Weeks 1-3), one-week post-therapy A (Week 7), one-week post-therapy B (Week 10), and four weeks after completing both therapies (Week 13)). Details of the design, procedure (recruitment process, therapy setup etc.), stimuli and data extraction (including fidelity check) were provided in Chapter 5.

6.2.2 Participants and clinical observations

Five adult participants with spoken word-finding difficulties post-stroke were recruited (see Chapter 5 for details, including our inclusion/exclusion criteria, recruitment process; Table 5.2 summarises each participant's demographic and relevant medical information). Scores for their baseline language and cognitive assessments are presented in Tables 6.1 and 6.2.

Table 6.1. Participants' scores on background language assessments

Assessment	<i>n</i> of items	P1	P2	P3	P4	P5
Spoken language processing						
PALPA 23. Spoken letter sound-written letter matching	26	24 (92.31%) *	25 (96.15%)	22 (84.62%) *	24 (92.31%) *	11 (42.31%) ***
PALPA 14a. Rhyme judgement: Picture selection (Rhyme) [†]	20	14 (70%)	7 (35%) ***	7 (35%) ***	5 (25%) ***	3 (15%) ***
PALPA 14b. Rhyme judgement: Picture selection (Non-rhyme) [†]	20	8 (40%) **	8 (40%) **	11 (55%)	13 (65%)	4 (20%) ***
CAT 12. Repetition of words	16 (Max score: 32)	28 (87.5%) *	29 (90.63%)	27 (84.38%) **	26 ^{a1} (81.25%) **	29 (90.63%)
CAT 13. Repetition of complex words	3 (Max score: 6)	6 (100%)	6 (100%)	5 (83.33%)	4 (66.67%) *	4 (66.67%) *
PALPA 8. Nonword repetition task [†]	30	18 (60%) **	21 (70%)	6 (20%) ***	8 (26.67%) ***	14 (46.67%) ***
Written language processing						
PALPA 18. Letter discrimination: Mirror reversal	36	36 (100%)	36 (100%)	36 (100%)	36 (100%)	36 (100%)
PALPA 22. Letter sounding [†]	26	4 (15.38%) ***	16 (61.54%) **	17 (65.38%) **	20 (76.92%)	2 (7.69%) ***
PALPA 22. Letter naming	26	24 (92.31%) *	18 (69.23%) ***	6 (23.08%) ***	26 (100%)	15 (57.69%) ***

PALPA 25. Visual lexical decision task	120	92 (76.67%) ***	96 (80%) ***	112 ^{a2} (93.33%) *	101 (84.17%) **	112 ^{a2} (93.33%) *
PALPA 35a. Spelling-sound regularity reading task: Regular words	30	15 (50%) ***	28 ^{a3} (93.33%)	17 (56.67%) ***	18 (60%) ***	22 (73.33%) ***
PALPA 35b. Spelling-sound regularity reading task: Exception, i.e., irregular, words	30	19 (63.33%) ***	26 ^{a3} (86.67%) *	19 (63.33%) ***	20 (66.67%) ***	23 (76.67%) ***
PALPA 36. Reading nonwords	24	0 (0%) ***	1 *** (terminated after 12 items)	6 (25%) ***	10 (41.67%) ***	1 (4.17%) ***

Semantic processing

Pyramid and Palm Trees: Three pictures version	52	49 (94.23%)	49 (94.23%)	49 (94.23%)	51 (98.08%)	44 (84.62%) ***
PALPA 47: Spoken word-picture matching (Max score: 40)	40	40 (100%)	38 (95%)	37 ^{a4} (92.5%) *	40 (100%)	35 (87.5%) ***
PALPA 48: Written word-picture matching (Max score: 40)	40	39 (97.5%)	39 (97.5%)	40 (100%)	40 (100%)	32 (80%) ***
PALPA 50. Written synonym judgment [†]	60	44 (73.33%) ***	55 (91.67%)	48 (80%) ***	52 (86.67%) **	50 (83.33%) ***

Naming

Boston Naming Test (BNT)	60	26 (43.33%) ***	28 (46.67%) ***	34 (56.67%) ***	21 (35%) ***	23 (38.33%) ***

	(Max score: 60)					
CAT 25. Writing picture names	5	4 (19.05%) ***	1 (4.76%) ***	17 (80.95%)	8 (38.10%) ***	14 (66.67%) **
	(Max score: 21)					

Endnotes for Table 6.1

Both raw scores and percentages (in parentheses) are presented in the table.

* Borderline / Mild impairment: Between 1.5 to 2 *SDs* below mean OR within 1 *SD* below the cut-off score (if cut-off is provided, e.g., for CAT).

This estimate is conservative and broadly follows Palmer, Boone, Lesser, & Wohl (1998), which diagnoses 1.3 *SDs* below mean as borderline and 2 *SDs* below mean as impaired.

** Moderate impairment: Between 2 to 3 *SDs* below mean or the cut-off score

*** Severe impairment: More than 3 *SDs* below mean or the cut-off score

^{a1} The bandwidth for the severity in the norms is very small due to the small *SD*. According to the norms, P4's performance on word repetition (CAT 12) is severely impaired. To consider P4's performance as "severely impaired" is too harsh; specifically, he was penalised for (a) requesting repetitions twice and (b) he made two errors. A 'moderate' severity rating was thus indicated.

^{a2} According to the norms, both P3 and P5 were severely impaired on PALPA 25 (Visual lexical decision task) for rejecting nonwords (53/60) but within normal limits for accepting real words (59/60). On balance, their performances were indicative of a mild impairment (or even potentially borderline within normal limits) because their individual overall scores (112/120; 93.33%) were relatively high compared to other participants as well as based on the author's clinical experience. Also, based on the author's understanding, PALPA norms were collected from young university students, which resulted in the high averages and small standard deviations for most subtests, which could be punitive when applied to the general public, especially older adults.

^{a3} The small *SD* means that a score of '28' out of 30 would be categorised as severely impaired. This is too punitive. Noting that only two errors were made, this is taken as grossly functional and for a score of '26', mildly-impaired. These scores are relatively high compared to other participants as well as patients the author had previously tested on the same task.

^{a4} The small *SD* means that although a score of '38' is within normal limits, a score of '37' would be categorised as moderately-impaired. Noting that only three errors were made out of forty items, this is taken as mildly-impaired.

[†] The norms for PALPA 8 (Nonword repetition), PALPA 14 (Rhyme judgement using picture selection), PALPA 22 (Letter sounding), and PALPA 50 (Visual synonym judgment) were not published in the official PALPA battery.

- For PALPA 8 and 22, there are no published norms. Provisional norms from a small sample of healthy UK native adults were collected ($n = 5$ Caucasians with English as L1, mean age = 60.8 ($SD = 10.71$)). Their performances were as follow: For PALPA 8 (Mean score: 25.4; $SD = 2.97$), and for PALPA 22 (Mean score = 21.8, $SD = 1.92$; a score below '16 was taken as indicating 'severely-impaired').
- For PALPA 14, the norms for typically-ageing adults ($n = 20$) reported by Harley, Oliver, Jessiman, & MacAndrew (2013) were used. To the author's knowledge, this is the only set of norms available for the test.
- For PALPA 50, the norms ($n = 21$) reported by Nickels and Cole-Virtue (2004) were used. Nickels and Cole-Virtue's norming samples for PALPA 50 were twice larger than another set of norms for the same test reported by Nightingale (1993; $n = 10$).

Table 6.2. Participants' scores on the background cognitive assessments

Assessment	Max score	P1	P2	P3	P4	P5
Attention						
Test of Everyday Attention (TEA)	7	6 (85.71%)	5 (71.43%) *	5 (71.43%) *	5 (71.43%) *	4 (57.14%) *
Subtest 2: Elevator counting						
TEA Subtest 3: Elevator counting with distraction	10	3 (30%) *	1 (10%) *	1 (10%) *	1 (10%) *	5 (50%) *
Memory						
CAT 4: Recognition memory	10	8 (80%)	8 (80%)	10 (100%)	8 (80%)	10 (100%)
Corsi Block tapping test	9	4 (44.44%)	4 (44.44%)	5 (55.56%)	6 (66.67%)	6 (66.67%)
Wechsler Memory Scale (WMS) 5a: Digits forward recall	8	4 (50%) *	3 (37.5%) *	4 (50%) *	3 (37.5%) *	3 (37.5%) *
WMS 5b: Digits backward recall	7	5 (71.43%)	3 (42.86%) *	2 (28.57%) *	3 (42.86%) *	3 (42.86%) *
Executive function and nonverbal reasoning						
Brixton Spatial Anticipation Test	55 (scored using number of errors)	34 (Accuracy: 21; 38.18%) *	23 (Accuracy: 32; 58.18%)	23 (Accuracy: 32; 58.18%)	35 (Accuracy: 20; 36.36%) *	33 (Accuracy: 22; 40%) *
Raven's Coloured Matrices	37	31 (83.78%)	35 (94.59%)	31 (83.78%)	27 (72.97%)	33 (89.19%)

* Impaired based on guidelines stated in the manuals. No guidelines or manualised norms were available for Corsi block tapping and Raven's Coloured Matrices. For Corsi block tapping, norms reported by Smirni, Villardita, and Zappalà (1983, *N* = 83 adults) were used, with the performance of the bottom 25th percentile taken as our cut-off (note that Smirni et al. did not provide data differentiated by age groups). For Raven's Coloured Matrices, norms reported by Measso et al. (1993, *N* = 894 adults) were used, with the mean for the age group taken to be the cut-off.

6.2.2.1 P1's clinical profile at baseline and probable locus/loci of impairment

The profiles of the participants are discussed in turn, with a focus the processes involved in word reading and spoken picture naming. Amongst the five participants, P1 appeared to have specific difficulties with visual word/nonword processing. He performed well on letter discrimination (PALPA 18), but was severely impaired on visual lexical decision (PALPA 25) and across reading tasks, including the reading aloud of regular words (PALPA 35a) and nonwords (PALPA 36), with these scores falling in the lower range of the sample's (Table 6.1). It was therefore likely that his primary loci of impairment resided in the (1) orthographic input lexicon and (2) grapheme-to-phoneme conversion pathway (Figure 6.1), which meant that all three of his orthographic input processing pathways (paths A, B and C) were impaired. This was also consistent with P1's self-report that he found reading simple text very difficult, despite enjoying text-heavy books in the past. Since his stroke, he reported preferring to look at pictures and relying on text-to-speech software to manage written text.

Specific to picture-naming, P1's conceptual and lexical-semantics were intact. This could be deduced from his intact performances on Pyramid and Palm Trees and both versions of word-to-picture matching tasks (PALPA 47/48; Table 6.1). However, he appeared to have post-semantic impairments, which in turn affected his spoken production. P1's performance on picture rhyme judgement was impaired due to high rates of false positives (tended to accept items as rhyming; PALPA 14b). Given that his performances on both word repetition tasks were within-normal-limits (CAT 12 and 13) in the context of impaired nonword repetition (PALPA 8), it was likely that his impaired performance on rhyme judgement was due to an impaired link between semantics and phonological output lexicon (on path 'A' in Figure 6.1). This could also explain why about 40% of P1's errors on Boston Naming Test (BNT) were circumlocutions (e.g., "fancy dress ... face ... teeth" for 'mask'), suggesting he recognised those pictures but had specific difficulty accessing the spoken forms stored in the phonological output lexicon. In addition, a proportion of his BNT naming errors were phonological errors (e.g., 'stelephone' for 'stethoscope'). Although intact on word repetition (CAT 12), he did make two errors and both were phonological (/trusifiks/ for 'crucifix' and 'gravity' for 'brevity'). Combined, these suggested mild deficits were present at the phonological output buffer. In essence, P1's primary loci of impairment resided in his (1) orthographic input lexicon, (2) sublexical grapheme-to-phoneme conversion pathway, and (3) post-semantic processes, specifically in the link between semantics and phonological output lexicon, as well as a mild deficit in his phonological output buffer.

6.2.2.2 P2's clinical profile at baseline and probable locus/loci of impairment

In contrast to P1, P2's reading aloud of regular and exception words were relatively intact (accuracy > 86% on PALPA 35 for reading aloud on both sets of words). Her performance across all semantic tasks were also strong (Table 6.1). However, P2's performance on visual lexicon decision (PALPA 25) was impaired and her picture-naming performance suggested an impaired post-semantic access to the phonological output lexicon (see below). Nonetheless, P2 likely relied on the combined activation drawn from the residual capabilities within lexical-semantic and direct lexical pathways (i.e., *both* paths A and B) to read aloud real words, since neither her performances in visual lexical decision nor picture naming were at floor; more importantly, she did successfully read aloud majority of the regular and exception words independently.

As for the sublexical route (path C), her impaired performances on both letter sounding (e.g., responses largely unrelated to target, like, /s3:/ for 't') and nonword reading (mostly null responses) suggested that her sublexical ability to convert orthography to phonology was severely compromised. Related to this was P2's performance on the 'Corsi Block Tapping Test', which assessed one's visual short-term memory span. She obtained the lowest score within the sample. Visual short-term memory had been linked to performance on nonword reading (see Predovan et al., 2009, for EEG evidence that linked visual short-term memory to nonword reading). P2 found nonword reading stressful, so testing was terminated after 12 items. On those 12 items, P2 could produce the beginnings of nonwords correctly most of the time (accuracy: 9/12; e.g., /k/ for 'ked', and /sn/ for 'snite'), but was not able to blend sounds together to produce the entire nonword. This was in spite of her relatively intact word reading, suggesting her ability to assemble novel sound sequences for articulation could be impaired.

In terms of spoken output processes relevant for picture-naming, P2's phonological output buffer was likely unimpaired, based on intact performances on word (CAT 12 and 13) and nonword repetitions (PALPA 8), as well as the absence of phonological errors during BNT picture naming. However, like P1, P2 appeared to have impaired post-semantic access to phonological output lexicon (part of path A). This was because: (a) P2's performance on rhyme judgement using pictures (PALPA 14) was similarly impaired, yet she demonstrated successful word repetition (CAT 12 and 13). The latter combined with her relatively intact reading aloud of both regular and irregular words, ruled out the possibility that her phonological output lexicon was impaired. (b) Also, although the majority of P2's naming errors were null responses (75%), it is worth noting that circumlocutions (e.g., "it's in Egypt" for 'pyramid') and rejections ("sword, not sword" for 'saw') made up half of her remaining naming errors. Thus, the profile across tasks suggested that P2's primary impairments were in the: (1) Post-semantic link to the phonological output lexicon (part of path A) and (2) sublexical pathway (path C).

6.2.2.3 P3's clinical profile at baseline and probable locus/loci of impairment

For reading aloud, P3's primary impairments appeared to reside partly in the lexical-semantic and direct lexical pathways between orthographic input lexicon and phonological output lexicon (paths A and B, Figure 6.1). Her direct lexical pathway (path B) was likely impaired: 20% of the errors she committed when reading exception words (PALPA 35) were semantic errors (e.g., 'married' for 'wedding' and 'dog' for 'wolf'), suggesting she could be relying on the lexical-semantic pathway (path A) to compensate for an impaired direct lexical route. However, P3's lexical-semantic path was, to some degree, also likely impaired as indicated by her semantic errors during reading (PALPA 35) and impaired performance on the written synonym judgment (PALPA 50). In addition, like both P1 and P2, P3 also appeared to have an impaired post-semantic access to her phonological output lexicon (part of path A). This was suggested by her impaired performance on rhyme judgment based on pictures (PALPA 14), despite having largely intact semantic systems (within normal limits on 'Pyramid and Palm Trees' and 'PALPA 48: Written Word-to-Picture Matching' task) and a functional phonological output lexicon. Like P1, P3's repetition of complex words appeared unimpaired (phonological output lexicon common to both paths D and E, Figure 4.1), but her ability to repeat nonwords was impaired (path F in Figure 4.1), suggesting that her phonological output lexicon should be relatively intact to enable complex word repetition. Lastly, her sublexical pathway (path C) was likely compromised, based on her impaired performances on letter sounding (PALPA 22) and nonword reading (PALPA 36).

Specific to spoken output processes relevant for picture naming, aside from the impaired post-semantic access to the phonological output lexicon (part of path A), P3 appeared to have deficits in the phonological output buffer. This could be deduced from P3's occasional phonological errors across tasks ($\approx 13\%$ of errors in BNT, e.g., /spinks/ for 'sphinx', both errors in word repetition were phonological errors, e.g., /prezident/ for 'president'). Thus, P3's main loci of impairments included: (1) Impaired access to the phonological output lexicon from semantics and from the orthographic input lexicon (part of paths A and B), (2) impaired sublexical processing (path C), as well as (3) a mild deficit in the phonological output buffer.

6.2.2.4 P4's clinical profile at baseline and probable locus/loci of impairment

P4's impairments in the processes related to word reading appeared distributed, although his abstract letter identification was intact (100% correct on 'PALPA 18: Letter Discrimination'), and he performed within normal limits converting graphemes to phonemes

(76.92% on PALPA 22: Letter Sounding; path C, Figure 6.1). His conceptual and lexical semantic systems appeared largely functional (100% correct on ‘Pyramid and Palm Trees’ and ‘PALPA 48: Written Word-to-Picture Matching’, although impairments (83.33%) were indicated by the more difficult ‘PALPA 50: Written Synonym Judgement’). Some degree of impairment appeared to be present in both his orthographic input lexicon (PALPA 35: Visual Lexical Decision; affecting both paths A and B) and phonological output lexicon (poor correct identification of rhymes in PALPA 14: Rhyme Judgment; also part of paths A and B). Although impaired, his nonword reading was the strongest within the sample (PALPA 36; path C). It would be interesting to see whether P4 could draw on the residual processes within paths A to C to benefit from the written cues.

However, one prominent locus of P4’s impairment affecting his spoken output would perhaps be the phonological output buffer, given his relatively large proportion of phonological errors during picture naming, word reading and word repetition (Phonological errors in: BNT $\approx$ 23%, e.g. /saipod/ for ‘tripod’, ‘harmonic’ for ‘harmonica’; PALPA 35 Word Reading: $\approx$ 27%, e.g., ‘pub’ for ‘pump’, /enərvɪ/ for ‘envy’; ‘CAT 12 and 13: Repetition of Simple and Complex Words’: 100%, e.g., /comfin/ for ‘conforming’), as well as the presence of a slight length effect (e.g., ‘CAT 12 Repetition of Word’: One-syllable items: 100% correct; three-syllable items: 75% correct). Overall, P4’s deficits appeared distributed (i.e., all three paths impaired to some extent), with a moderate deficit in the phonological output buffer.

6.2.2.5 P5’s clinical profile at baseline and probable locus/loci of impairment

Finally, P5 was the only participant with impaired conceptual and lexical-semantic systems (see performances on ‘Pyramid and Palm Trees’ and ‘PALPA 47/48: Spoken and Written Word-to-Picture Matching’ in Table 6.1; i.e., path A in Figure 6.1 was impaired). This could explain her relatively high proportion of semantic errors during picture naming (BNT: Semantic errors $\approx$ 31.8%, e.g., ‘mattress’ for ‘hammock’). Her letter-to-sound conversion pathway (path C) also appeared to be severely impaired (see ‘PALPA 22: Letter Sounding’, ‘PALPA 36: Reading Nonwords’ in Table 6.1).

The rest of P5’s impairments appeared distributed, and of lesser severity, in the other processing components: Specifically, direct lexical reading via path B was probably impaired, as P5 was unable to entirely compensate for her compromised sublexical and semantic-lexical pathways, by relying on path B when reading regular and irregular words (e.g., “vroom ... tat, tat, the the work” for ‘routine’, an exception word). Her phonological output buffer was also likely impaired as she showed, for example, length effect in ‘PALPA Syllable Length Reading’ (One-syllable items: 100%, two-syllable items: 75%, three-syllable items: 62.5%), and made phonological errors during word reading and nonword repetition

(e.g., 'apium' for 'opium', 'smog' for 'smode'). Therefore, P5's reading was impaired to varying degree within the orthographic processing pathways A to C (Figure 6.1). However, her primary loci of impairment would be: (1) impaired semantics (within path A), which certainly affected her naming, as well as (2) an impaired ability to convert graphemes to phonemes (path C).

Overall, none of participants appeared to be intact on all three processing pathways (paths A to C, Figure 6.1). P2 appeared to be least severely impaired, but she too was impaired on written nonword processing (path C). Compared to the others, P1 appeared to have a relative deficit in orthographic *input* processing (e.g., his performance on visual lexical decision was the weakest amongst the participants). He consequently struggled with reading written stimuli, though was still able to read aloud 50% of the regular words (Table 6.1). For the other participants, their primary impairments related to visual word reading appeared to occur after the orthographic input lexicon (e.g., post-semantic lexical access for P3 or at the semantic level, for P5). Each participant in turn capitalised on their residual capabilities differently to read aloud the orthographic forms during baseline assessments.

6.2.3 Individual analyses for primary research questions (1) and (2): Outcomes of cue conditions across and between time points

Item-level data were analysed per participant for each of the five conditions (recapped below, but see section 5.2.3.2 on details of the conditions) using WEST. Step-by-step procedure of WEST can be found in Howard and colleagues (2015); related discussion on its advantages, including how it accounts for autocorrelations, can be found in Chapter 4 (p. 115). As a reminder, the five conditions were:

- Condition (1): Initial Letter Cue: Orthography only
- Condition (2): Repeated Exposure: Initial Letter Cue (i.e., Untreated items encountered during 'Initial Letter' orthographic cue therapy)
- Condition (3): Whole Word Cue: Orthography only
- Condition (4): Repeated Exposure: Whole Word Cue (i.e., Untreated items encountered during 'Whole Word' orthography cue therapy), and
- Condition (5): Unseen Control (i.e., Items only encountered during assessments but not during therapy)

WEST-Trend and ROC analyses were conducted, with scores at each time point adjusted by different weights computed based on Howard and colleagues. To recap,

picture-naming data was collected on six occasions, i.e., three baseline sessions, one-week after therapy 1 (Post-therapy assessment 1 on Week 7), one-week after therapy 2 (Post-therapy assessment 2, Week 10) and four-weeks after the completion of both therapies (Post-therapy assessment 3, Week 13). The conditions delivered first (i.e., whichever therapy cue and its attendant repeated exposure condition) were assigned the following set of weights (-5, -3, -1, 1, 3, 5) and (5, -4, -13, 13, 4, -5) for WEST-Trend and ROC respectively. Conditions delivered second were assigned (-5, -3, -1, 1, 3, 5) and (25, 1, -23, -47, 34, 10) as corresponding weights. Data for the control condition (untreated, unexposed) were assigned the same weights as the conditions delivered first, to account for transfer effects that may arise starting from the first therapy. For each condition, the weighted results were then tested using a one-tailed one-sample *t*-test. Following Howard and colleagues' guidelines, findings from a condition was considered significant only if *both* WEST-Trend and WEST-ROC results were significant. Similar to Chapter 3, as follow-up, item-level post-hoc analyses were pursued using McNemar for conditions with significant findings pertaining to binary accuracy outcomes, and Wilcoxon signed ranks test for continuous RT outcomes.

6.2.4 Individual-level exploratory analyses for secondary research question (3): Associations between performances on baseline assessments and therapy outcomes

As the goal is to identify abilities involved in written cue processing, only language tasks pertaining to orthographic decoding will be relevant (see Figure 6.5 below for the tasks and the pathways involved). All cognitive tasks administered at baseline were also included (Figure 6.5). For each task, a group mean and standard deviation were calculated based on the pooled score across all five participants. For every participant, a z-score was then computed for each task. Finally, on each task, participants were arranged according to their z-scores in ascending order, with the group median being the middle score.

Next, participants were divided into groups according to the therapy cue(s) that was/were effective for them, i.e., the unit of written cue(s) that significantly improved naming on any main outcome measure for the participant. In terms of analysis, a systematic pattern is present if there is a consistent pairing between the group type and task performance. Also, if multiple participants significantly improved following one particular unit of cue, then it would be interesting to find out the common language/cognitive task in which all members within that group excelled, relative to non-members (i.e., z-scores on a task are higher relative to non-members). A high z-score is defined as a value found above the median score.

The reverse also applies. A systematic pattern is present if a lack of ability to benefit from a cue unit is consistently tied to a deficit in language/cognitive ability (i.e., z-scores on a task are lower relative to others). A low z-score would be one that falls below the median.

6.3 Results

6.3.1 Primary research questions (1) and (2): At the individual participant level, which unit of the orthographic cue will result in significant improvement; also, how durable are the therapy effects of each orthographic cue?

Descriptive statistics for the three main outcome measures for each individual are shown in Figures 6.2 to 6.4 below. Table 6.3 summarises the results from group- and individual-level WEST and post-hoc analyses. The group-level results in this table are excerpted from Table 5.6. Significant findings are highlighted in yellow to assist reading. Detailed statistical results for each participant are published in the appendices (see Appendix 6.1 for the complete WEST-Trend or ROC findings, and Appendix 6.2 for related post-hoc follow-ups, conducted where applicable).

The key results are described in the sub-sections below.

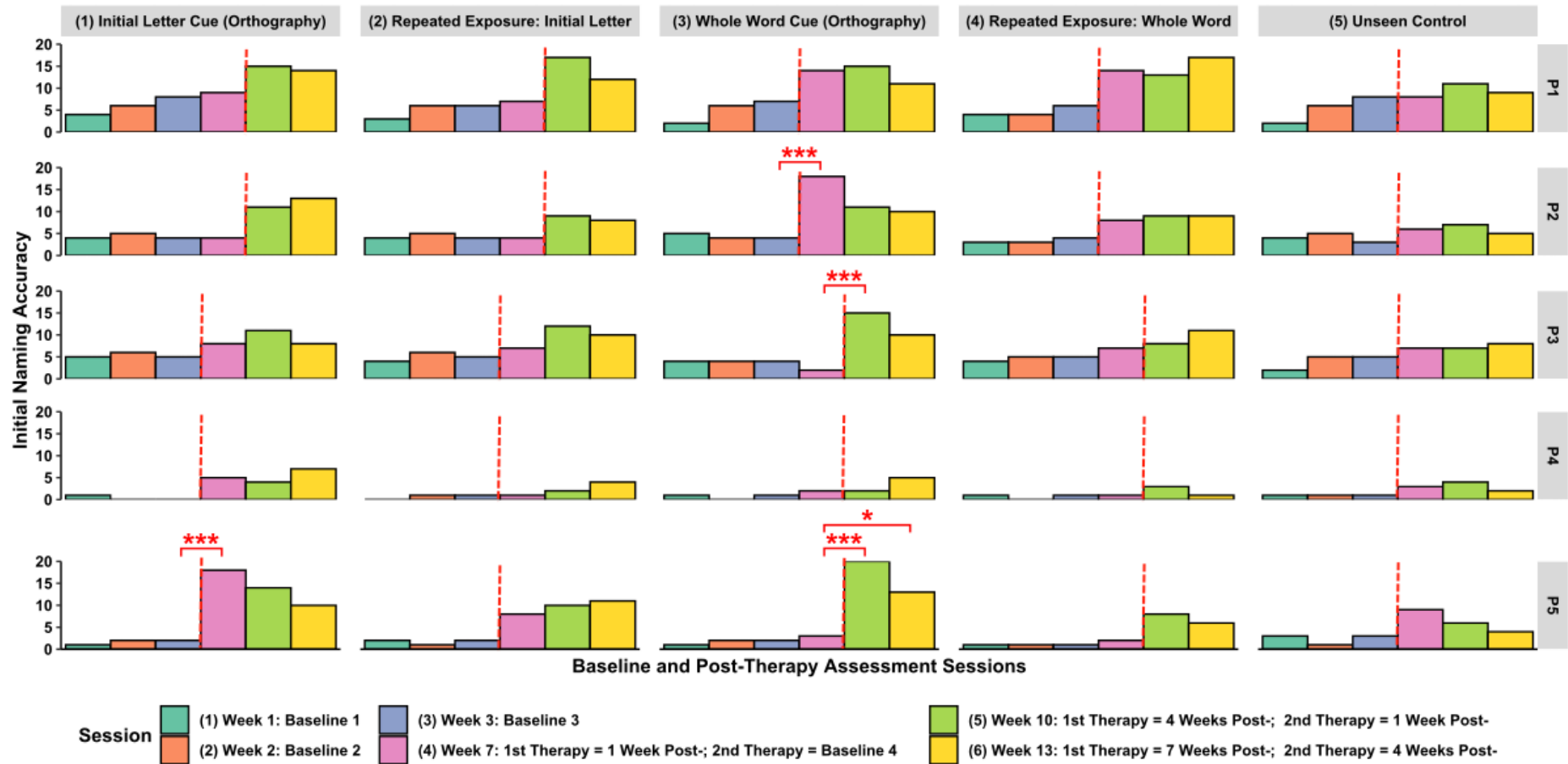
6.3.1.1 Initial letter orthographic cue therapy

Based on WEST-Trend and ROC, three (P2, P4, and P5) out of the five participants showed significant naming improvement across the six baseline and post-therapy assessment sessions on at least two main naming outcome measures, following the initial letter orthographic cue therapy (Table 6.3).

In the first set of post hoc comparisons whereby results from post-therapy assessments at different time-points were pitted against that of the final baseline, significant results were observed for all three participants (P2, P4 and P5) at one and four-weeks post-therapy on at least two main measures of naming outcomes [ps at least $< .04995$; Appendix 6.2]. Improvements were also significantly better than baseline when assessed seven weeks afterwards for P4 and P5 on all three main naming outcomes [for P4: $ps < .02$; for P5: $ps < .011$; Appendix 6.2].

Figure 6.2.

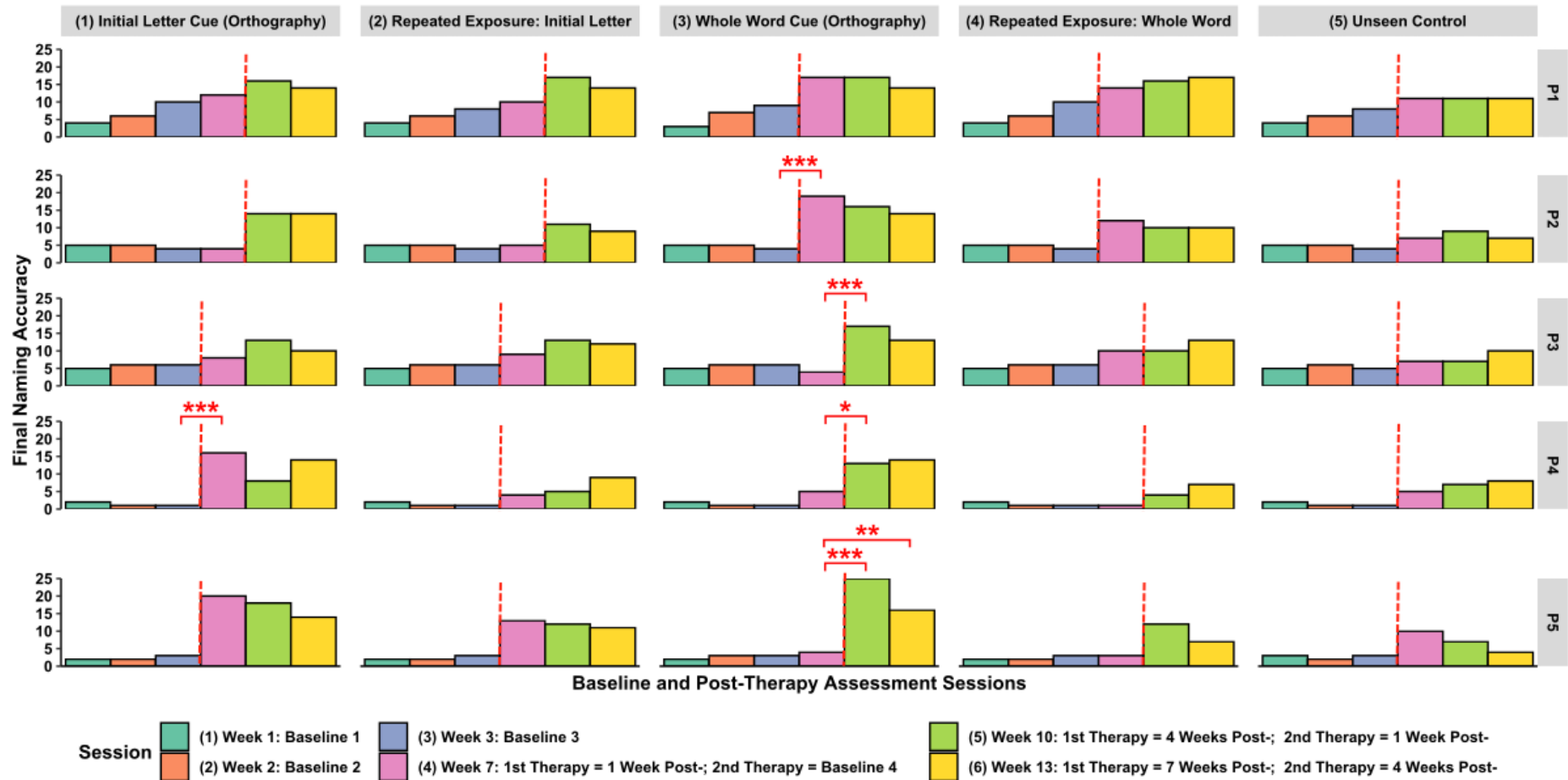
Initial naming accuracy scored for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated.



For post hoc - Conditions with four baselines (adjusting for two comparisons using Bonferroni's): + .025 < p < .05, * .01 < p < .025, ** .001 < p < .01, *** p < .001
 For post hoc - Conditions with three baselines (adjusting for three comparisons using Bonferroni's): + .01667 < p < .05, * .01 < p < .01667, ** .001 < p < .01, *** p < .001

Figure 6.3.

Final naming accuracy scored for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated.

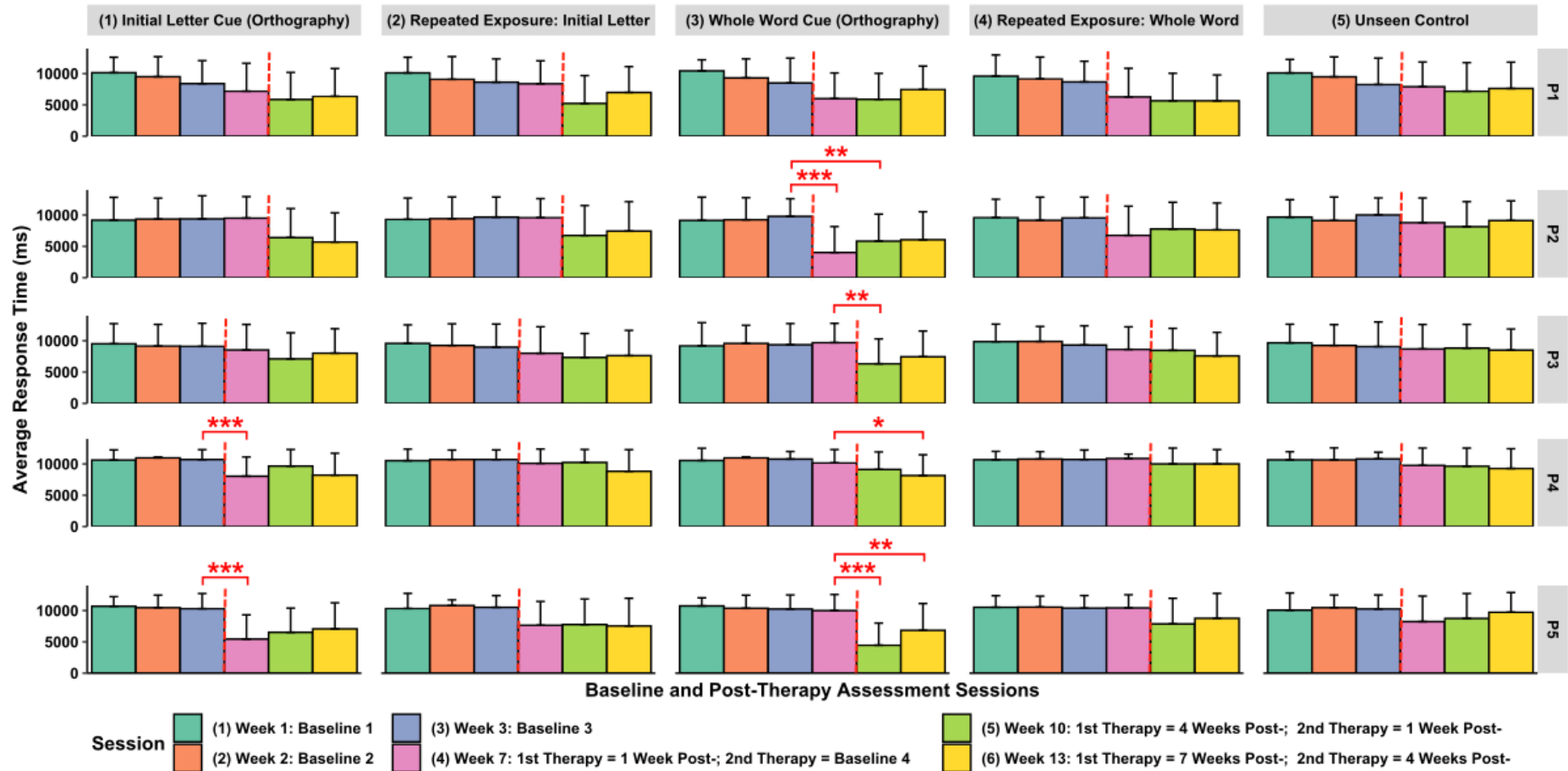


For post hoc - Conditions with four baselines (adjusting for two comparisons using Bonferroni's): + .025 < p < .05, * .01 < p < .025, ** .001 < p < .01, *** p < .001

For post hoc - Conditions with three baselines (adjusting for three comparisons using Bonferroni's): + .01667 < p < .05, * .01 < p < .01667, ** .001 < p < .01, *** p < .001

Figure 6.4.

Average RT of correct responses for each participant across the six assessment sessions in the two orthographic cues, repeated exposure and control condition respectively (with the dotted red line inserted between final baseline and the first post-therapy assessment). For post hoc findings, only results that remained significant after comparison against repeated exposure data are indicated.



For post hoc - Conditions with four baselines (adjusting for two comparisons using Bonferroni's): + .025 < p < .05, * .01 < p < .025, ** .001 < p < .01, *** p < .001

For post hoc - Conditions with three baselines (adjusting for three comparisons using Bonferroni's): + .01667 < p < .05, * .01 < p < .01667, ** .001 < p < .01, *** p < .001

Table 6.3. Summary of therapy findings for each condition for group and individual-level (WEST and nonparametric post-hoc) analyses for each main outcome measure. Group-level results are a reproduction of Table 5.6.

Condition		Initial Letter Cue (Ortho)			Repeated Exposure: Initial Letter Cue (Ortho)			Whole Word Cue (Ortho)			Repeated Exposure: Whole Word Cue (Ortho)			Unseen Control: Untreated Unexposed			
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	
Group (By Pt)		*	*	*	*	*	*	*	*	*	*	*	*	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	
Post-hoc (against final baseline)	x1 Week. Post- vs. BL	*	*	*	†	*	*	†	*	*	*	*	†	*	—	—	—
	x4 Weeks Post- vs. BL	*	*	*	*	*	*	*	*	*	*	†	*	*	—	—	—
Post-hoc (against rep. exposure)	Final Baselines	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				—	—	—	
	x1 Week Post-	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				<i>n.S.</i>	*	*							
x4 Weeks Post	<i>n.S.</i>	<i>n.S.</i>	*	<i>n.S.</i>				†	<i>n.S.</i>								
		<i>(p = .0796)</i>						<i>(p = .0796)</i>									
P1 (W-Trend)		***	***	***	***	***	***	***	***	***	***	***	***	***	**	***	
P1 (W-ROC)		<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	*	†	*	*	*	*	*	<i>n.S.</i>	*	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	
Post-hoc (against final baseline)	x1 Week. Post- vs. BL	—	—	—	**	*	**	<i>n.S.</i>	†	†	†	—	**	—	—	—	
	x4 Weeks Post- vs. BL	—	—	—	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	†	**	**	*	—	**	—	—	—	
Post-hoc (against rep. exposure)	Final Baselines	—	—	—				<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				—	—	—	
	x1 Week Post-	—	—	—				<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>							
x4 Weeks Post-	—	—	—	<i>n.S.</i>				<i>n.S.</i>	<i>n.S.</i>								
								<i>n.S.</i>									
P2 (W-Trend)		**	***	***	*	*	*	***	***	***	**	**	**	<i>n.S.</i>	†	<i>n.S.</i>	
P2 (W-ROC)		**	***	**	†	*	**	***	***	***	<i>n.S.</i>	**	*	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	
Post-hoc (against final baseline)	x1 Week Post- vs. BL	*	**	**	<i>n.S.</i>	†	**	***	***	***	—	**	**	—	—	—	
	x4 Weeks Post- vs. BL	**	**	**	<i>n.S.</i>	<i>n.S.</i>	†	<i>n.S.</i>	**	**	—	†	†	—	—	—	
Post-hoc (against rep. exposure)	Final Baselines	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				<i>n.S.</i>	<i>n.S.</i>	†				—	—	—	
	x1 Week Post-	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				*	*	*							
x4 Weeks Post-	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>	<i>n.S.</i>				<i>n.S.</i>	†								
								<i>n.S.</i>									

(continued)

Condition		<u>Initial Letter Cue (Ortho)</u>			<u>Repeated Exposure: Initial Letter Cue (Ortho)</u>			<u>Whole Word Cue (Ortho)</u>			<u>Repeated Exposure: Whole Word Cue (Ortho)</u>			<u>Unseen Control: Untreated Unexposed</u>		
		IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT	IN Acc.	FN Acc.	RT
P3 (W-Trend)		†	*	*	**	***	**	***	***	**	**	**	**	**	*	*
P3 (W-ROC)		<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	***	***	**	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against final baseline)	x1 Week Post- vs. BL	—	—	—	—	—	—	***	***	**	—	—	—	—	—	—
	x4 Weeks Post- vs. BL	—	—	—	—	—	—	†	†	<i>n.s.</i>	—	—	—	—	—	—
	x7 Weeks Post- vs. BL	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Post-hoc (against rep. exposure)	Final Baselines	—	—	—	—	—	—	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	—	—
	x1 Week Post-	—	—	—	—	—	—	†	†	*	—	—	—	—	—	—
	x4 Weeks Post-	—	—	—	—	—	—	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	—	—
	x7 Weeks Post-	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P4 (W-Trend)		**	***	***	*	**	**	*	***	***	<i>n.s.</i>	*	*	<i>n.s.</i>	**	**
P4 (W-ROC)		*	***	***	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	**	*	<i>n.s.</i>	*	**	†	*	†
Post-hoc (against final baseline)	x1 Week Post- vs. BL	<i>n.s.</i>	***	***	—	—	—	—	*	<i>n.s.</i>	—	<i>n.s.</i>	<i>n.s.</i>	—	<i>n.s.</i>	<i>n.s.</i>
	x4 Weeks Post- vs. BL	<i>n.s.</i>	*	†	—	—	—	—	*	*	—	<i>n.s.</i>	<i>n.s.</i>	—	<i>n.s.</i>	†
	x7 Weeks Post- vs. BL	*	***	**	—	—	—	—	*	*	—	<i>n.s.</i>	<i>n.s.</i>	—	†	†
Post-hoc (against rep. exposure)	Final Baselines	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	—	—
	x1 Week Post-	<i>n.s.</i>	**	*	—	—	—	—	**	<i>n.s.</i>	—	—	—	—	—	—
	x4 Weeks Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	**	<i>n.s.</i>	—	—	—	—	—	—
	x7 Weeks Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	<i>n.s.</i>	*	—	—	—	—	—	—
P5 (W-Trend)		***	***	***	***	***	***	***	***	***	**	**	***	*	*	*
P5 (W-ROC)		***	***	***	*	***	***	***	***	***	*	**	**	*	**	**
Post-hoc (against final baseline)	x1 Week Post- vs. BL	***	***	***	<i>n.s.</i>	**	**	***	***	***	<i>n.s.</i>	**	**	<i>n.s.</i>	†	*
	x4 Weeks Post- vs. BL	***	***	***	†	**	**	*	**	**	<i>n.s.</i>	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
	x7 Weeks Post- vs. BL	*	**	*	*	†	*	—	—	—	—	—	—	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Post-hoc (against rep. exposure)	Final Baselines	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	—	—
	x1 Week Post-	*	<i>n.s.</i>	†	—	—	—	**	***	**	—	—	—	—	—	—
	x4 Weeks Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	†	*	†	—	—	—	—	—	—
	x7 Weeks Post-	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	—	—	—	—	—	—	—	—	—	—	—	—

Endnotes for Table 6.3:

IN Acc.: Initial Naming Accuracy, FN Acc.: Final Naming Accuracy, Pts: Participants, W-Trend: WEST-Trend, W-ROC: WEST-ROC
Ortho: Orthography

For the first four conditions: Group analyses based on the full sample ($n = 5$) were conducted across three time-points, i.e., the final baseline prior to cue therapy and the two subsequent post-therapy assessments (i.e., one week post- and four weeks post-therapy). For the 'Control: Untreated and Unexposed' condition, separate planned group analyses were run for participants whose first therapy was letter cue versus word cue respectively across the same three time-points; both sets of results were not significant across the three outcome variables. These are the same results published in Chapter 5, but included to facilitate comparison.

For group analyses, WEST analyses and post-hoc (against rep. exposure) comparisons: † $.05 < p < .07$, * $.01 < p < .05$, ** $.001 < p < .01$, *** $p < .001$

For post-hoc (against baseline just prior to therapy) - Conditions with four baselines (after adjusting for **two comparisons** using Bonferroni's):
† $.025 < p < .05$, * $.01 < p < .025$, ** $.001 < p < .01$, *** $p < .001$

For post-hoc (against baseline just prior to therapy) – Conditions with three baselines (after adjusting for **three comparisons** using Bonferroni's):
† $.01667 < p < .05$, * $.01 < p < .01667$, ** $.001 < p < .01$, *** $p < .001$

—: Indicates where it is not necessary to conduct post-hoc analyses, e.g., either preceding omnibus finding using combined WEST was not significant or no data were collected for that particular time-point

Yellow spaces indicate the results that are significant.

Grey spaces indicate areas where it does not make sense to conduct analyses, e.g., it is not possible to compare repeated exposure condition against itself

Table 6.4. Changes in the raw number of items correctly named and improvement in average RT (from baseline immediately prior to therapy), and their proportional changes¹ specified within parentheses. Only post hoc results that remained significant after comparison against repeated exposure data are reported.

Change in Performance in: Wk(s) Post- Therapy:	<u>Initial Letter Orthographic Cue</u>									<u>Whole Word Orthographic Cue</u>								
	<u>Initial Naming Accuracy</u>			<u>Final Naming Accuracy</u>			<u>Average RT (ms)</u>			<u>Initial Naming Accuracy</u>			<u>Final Naming Accuracy</u>			<u>Average RT (ms)</u>		
	1 Wk	4 Wks	7 Wks	1 Wk	4 Wks	7 Wks	1 Wk	4 Wks	7 Wks	1 Wk	4 Wks	7 Wks	1 Wk	4 Wks	7 Wks	1 Wk	4 Wks	7 Wks
P1 (<i>n</i> = 26)																		
P2 (<i>n</i> = 24)										14*** (0.70)			15*** (0.75)			5778.1 *** (0.59)	3974.5** (0.41)	
P3 (<i>n</i> = 25)										13*** (0.57)			13*** (0.62)			3382.6 ** (0.35)		
P4 (<i>n</i> = 28)				15*** (0.56)			2668.6 *** (0.25)						8* (0.35)				2026.4* (0.20)	
P5 (<i>n</i> = 28)	16*** (0.62)						4868.4 *** (0.47)			17*** (0.68)	10* (0.40)		21*** (0.88)	12** (0.50)		5517.3 *** (0.55)	3147.9** (0.32)	

For post-hoc (against baseline just prior to therapy) - Conditions with four baselines (after adjusting for **two comparisons** using Bonferroni's; see Figure 5.2 for number of baselines): † .025 < *p* < .05, * .01 < *p* < .025, ** .001 < *p* < .01, *** *p* < .001

For post-hoc (against baseline just prior to therapy) – Conditions with three baselines (after adjusting for **three comparisons** using Bonferroni's; see Figure 5.2 for number of baselines): † .01667 < *p* < .05, * .01 < *p* < .01667, ** .001 < *p* < .01, *** *p* < .001

¹ Proportional (propn) change follows the calculation specified in Section 3.2.4.6.2.

6.3.2.2 Whole word orthographic cue therapy

Significant overall improvement across time was obtained for the five participants, following the whole word written cue therapy on final naming accuracy and RT, according to WEST-ROC and Trend analyses (Table 6.3), but this needs to be qualified against the improved naming obtained from non-cue related effects as observed in their performance following repeated exposure reported below.

In post-hoc analyses, significant results were also obtained for the five participants post whole word cue therapy when compared against the respective final baselines, on at least one main measure of naming outcome at both one and four-weeks after therapy. P2 also maintained her significant naming improvement over baseline (final naming accuracy and RT) when assessed seven weeks afterwards [$p = .006$ for final naming accuracy; $p = .002$ for RT; Appendix 6.2]. Again, this set of results have to be qualified against the improvements stemming from repeated exposure (next section).

6.3.2.3 Comparing against improved outcomes from repeated exposure

According to WEST-Trend and ROC, four participants also improved based on repeated attempts at picture-naming in the absence of written cues (P1, P2, P4, and P5). When significant findings from cue conditions were compared against the corresponding significant results from repeated exposure conditions: (1) Only results for *whole word orthographic cue therapy* remained significant for three participants at both one and four-weeks post-therapy (P2, P4, and P5); (2) Results for *initial letter cue therapy* were significant for two participants, lasting up to one week but not four weeks post-therapy (P4 and P5). Results for P1, however, were no longer significant when his performance from whole word cue therapy were compared against the corresponding repeated exposure results.

Finally, repeated exposure conditions did not result in significant overall improvement over time for P3, but she demonstrated significant progress following whole word orthographic cue therapy. A follow-up verification to check how well her naming improvements from whole word cue therapy could stand up against any influence from repeated encounters was run: At one-week post-therapy, her speed of naming (RT) remained significantly faster over and beyond improvements stemming from repeated encounters [$p = .0196$]. In addition, she was also naming marginally more accurately, when compared against her performance induced by repeated exposure [$p = .065$ for both initial and final naming accuracies].

To summarise the results, the five participants could be divided into three groups based on the therapy cue(s) effective for them, after unrelated improvements from repeated exposure were taken addressed through direct comparisons. The groupings were:

- Group (1) – Benefitted from neither initial letter nor whole word orthographic cues (P1),
- Group (2) – Benefitted significantly only from the whole word orthographic cues (P2 and P3),
- Group (3) – Benefitted significantly from *both* initial letter and whole word orthographic cues (P4 and P5)

For readers interested in the amount of progress made in terms of raw values following the orthographic cues, the information for each individual (where applicable) is compiled in Table 6.4. To avoid confusing significant therapy effects from the improvements generated from repeated encounters with items, only data whose results remained significant after comparing against repeated exposure were included. It is worth noting that for the initial letter orthographic cue, the greatest amount of improvement was recorded by P5: Compared to her performance at the baseline prior to therapy, she improved by 16 items when assessed one week later, based on initial naming outcome (Figure 6.2). P5's RT also improved as much as 4868.4 ms in the same period (Figure 6.4). When directly evaluated against repeated exposure, all results for initial letter cues were not significant beyond one-week, and thus raw values were not enumerated further.

For the whole word orthographic cue, the greatest progress was also registered by P5, who improved by 17 and 21 items respectively based on initial and final naming outcomes, at one-week post-therapy. The smallest (but significant) change in accuracy was a respectable additional 8 items produced by P4, on his final naming outcome one week later. The largest improvement in average RT was 5778.1 ms obtained by P2, at one-week post-treatment. The largest improvement at four weeks post-treatment were an additional 12 items successfully named by P5 based on her final naming outcome, and a reduction of 3974.5 ms in latency by P2 (all changes graphically captured in Figures 6.2 to 6.4).

6.3.2.4 Unseen controls

Improvements on untreated and unexposed items were also observed for two participants across time, based on WEST Trend and ROC (P4 and P5). When compared against the final baseline prior to therapy, their improvements were marginal (Table 6.3;

Appendix 6.2), although P5's faster naming speed did remain significant [$p = .013$ on RT], albeit only so for a week after the first therapy (initial letter orthographic cue).

6.3.2.5 Control task: Nonword repetition

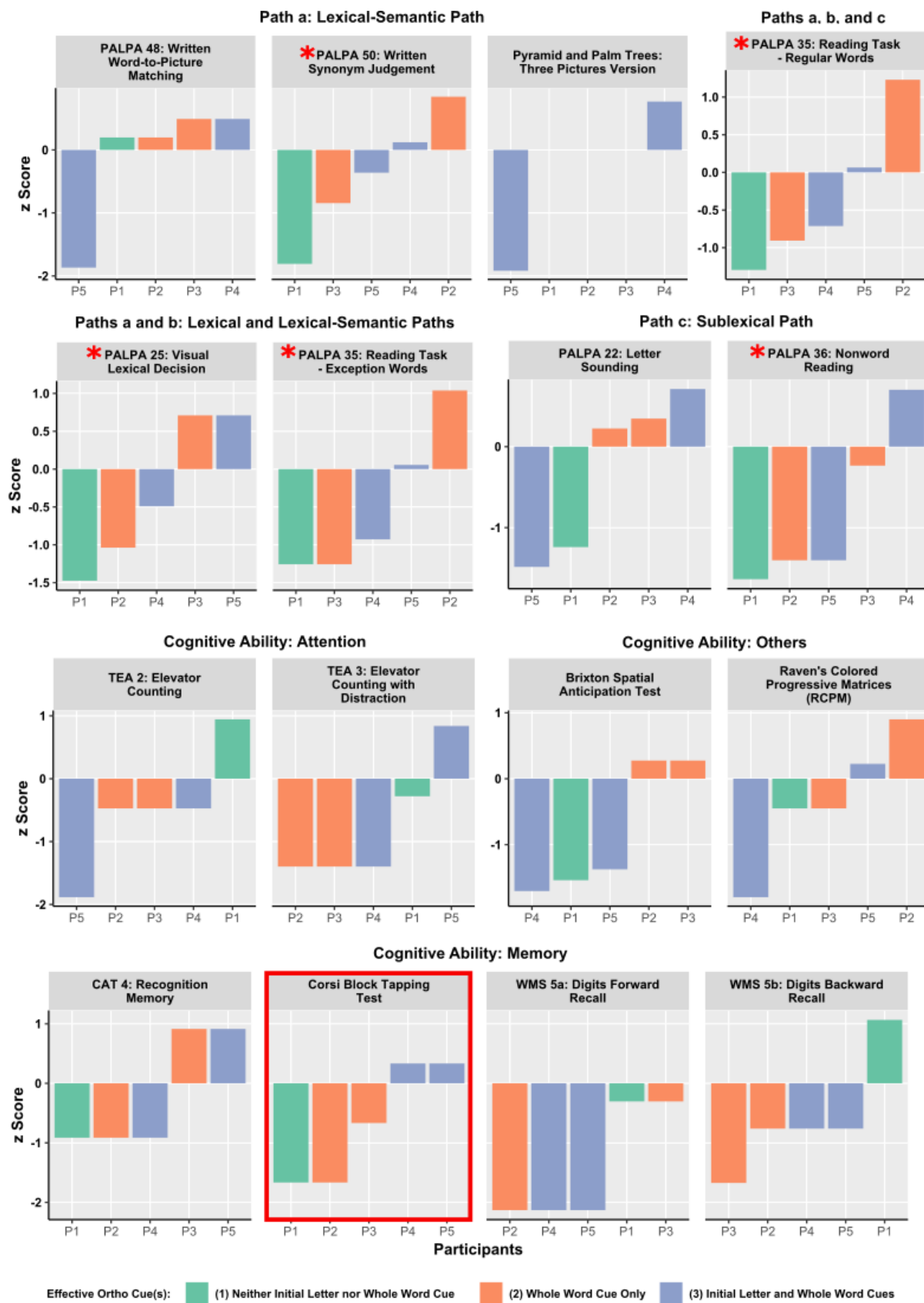
Finally, item-level analyses were conducted using Cochran's Q test for each participant's performance on nonword repetition across the four time points, i.e., at the final baseline (Week 3), one-week after the participant's first therapy (Week 7), one-week after the second therapy (Week 10), and at the maintenance check (Week 13). None of the participants' results was significant [P1: $\chi^2(3, N = 30) = 1.08, p = .78, \mathcal{R} = 0.015$; P2: $\chi^2(3, N = 30) = 0.72, p = .87, \mathcal{R} = 0.013$; P3: $\chi^2(3, N = 30) = 2.06, p = .56, \mathcal{R} = 0.0057$; P4: $\chi^2(3, N = 30) = 5.03, p = .17, \mathcal{R} = 0.011$; P5: $\chi^2(3, N = 30) = 0.56, p = .91, \mathcal{R} = 0.014$; Berry, Johnston and Mielke, 2007].

6.3.2 Secondary research question (3): What is the relationship between each participant's baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue?

Analyses from primary research questions suggested that the participants could be divided into three distinct groups, within which two systematic patterns could potentially be identified. Pattern 1 would help us distinguish between the individuals who did not benefit from either orthographic cue from those who could benefit from at least the whole word orthographic cue (i.e., Group (1) versus Groups (2) and (3)). The second pattern distinguished between those who could benefit from only the whole word written cues from individuals who benefitted from *both initial letter and whole word* orthographic cues (i.e., Group (2) versus Group (3)).

Results from the z-scores computation are displayed in Figure 6.5 below. The bars were arranged in ascending order, such that participants whose performance was relatively weaker than the rest of the sample on that task were placed on the left (i.e., task tapped a skill that is a relative deficit) and those stronger on the right (i.e., relative strength) to the median. Of interest are the systematic pairings between: (1) The group that found neither orthographic cues effective (i.e., green bars) and potential relative deficit(s) this group was associated with, as well as (2) The group that benefitted significantly from *both* initial letter and whole word orthographic cues (i.e., blue bars) and potential relative strength(s) this

Figure 6.5. Participants' z-scores on baseline language and cognitive assessments and their groupings according to the therapy cue(s) found effective. Data sorted based on z-scores.



group was associated with (i.e., looking at results consistently located on the right of the median across members within the group).

P1 was the only individual who benefitted from neither unit of orthographic cues. He consistently scored the lowest within the sample on a subset of language tasks marked with a red asterisk in Figure 6.5 (green bar consistently located on the left-most side of each of these graphs). The four tasks were orthographic input processing tasks at the level of the letter string, i.e., 'PALPA 25: Visual lexical decision', 'PALPA 35: Reading of regular and exception words', 'PALPA 36: Nonword reading', and 'PALPA 50: Written synonym judgement'. On these four tasks, performances for the rest of the participants were systematically stronger than P1, although their relative positions varied.

On the other hand, P4 and P5 represented the individuals who benefitted from *both* units of orthographic cues (blue bars). Based on Figure 6.5, there is only one task in which both P4 and P5 scored relatively higher than the rest of the sample, i.e., 'Corsi Block Tapping Test,' (data boxed within a red rectangle, with both blue bars to the right of the median). No further systematic pairings between group types and task performances could be discerned for the remaining baseline assessments.

An alternative layout of the results is plotted in Appendix 6.3 for readers who prefer their data on the x-axes organised according to the participants' group outcome (i.e., grouping based on 'effective therapy cue(s)').

6.4 Discussion

In chapter 5, evidence for the initial letter and whole word orthographic cues were provided using group analysis, and it was shown that the effects were significantly stronger than repeated exposure at one-week post-therapy and marginally so, four weeks afterwards. In this chapter, these group-level results were unpacked further, using individual analyses. There was an attempt to refine the evidence base by asking which participants benefitted from which unit of orthographic cue (Research question 1), and how long did the benefits last for each participant (Research question 2). Based on these results, participants were divided into three distinct groups according to the unit of written cues effective for them. Lastly, secondary research question (3) called for an exploration to identify common language and/or cognitive abilities that would help allocate participants into their distinct therapy cue groups. These common abilities might represent the underlying mechanism(s)

participants within the same group could be capitalising on, to engage with a particular unit of written cue for anomia rehabilitation.

Research question (1): Outcomes of each unit of orthographic cue and the implications of repeated exposure

Results from WEST-Trend and ROC analyses showed that all five participants demonstrated significant improvement following either the initial letter or whole word orthographic cue therapy. However, these improvements were not entirely cue-related because four of the participants (P1, P2, P4, and P5) also significantly improved based on repeated encounters with the pictures, in the absence of cueing (Appendices 6.1 and 6.2; or see Table 6.3 for a summary of group and individual-level findings). To ensure rigour, results from the cue conditions were directly compared against findings from corresponding repeated exposure conditions. Against repeated exposure, therapy results from whole word orthographic cues remained significant for four out of the five participants (P2, P3, P4, and P5) and the results were replicated across more than one outcome measure per participant, attesting to the whole word cues' robustness. For the initial letter cues, results from only two participants (P4 and P5) remained significant. Results from the control task, nonword repetition, were not significant for all five participants, suggesting that any gains observed could not be simply attributed to general improvement over time.

These results are crucial because: Firstly, they reiterated the importance of accounting for effects from repeated exposure (also see pp. 131-132 on a similar point, but from the vantage point of a facilitation study). This finding is also consistent with recent individual case series reported in literature (Creet et al., 2019). Notably, it was only through individual-level analyses in this chapter that it was possible to pinpoint which participants' improvement were likely attributed to non-therapy related reasons. In our sample, this would be P1, whose significant improvement following whole word cue therapy could not be reliably distinguished from the improvement he made based on repeated encounters (Table 6.3). Encouragingly, one has greater confidence that there were at least four participants (P2 to P5) whose improvement were not therapy-unrelated but could be clearly attributed to the orthographic cues.

Secondly, group-level findings in Chapter 5 showed that on average, both initial letter and whole word orthographic cues were effective. In this chapter, greater clarity is shed on this broad evidence by specifying exactly how many participants and exactly who benefitted from which written cue. Specific to the whole word written cue, there were four participants (P2 to P5, 80% of sample) who benefitted. Interestingly, the four participants did not share a

common aphasia diagnosis based on WAB, and their periods of time post-stroke varied substantially (from 10 months to more than 8 years; Table 5.2). These details suggest the versatility of the whole word orthographic cue to successfully rehabilitate a range of individuals with different profiles, as long as they can process written words. Three of these participants also improved on all three outcome measures (P2, P3 and P5), attesting to the cue's comprehensive impact. Case-level evidence from P2 to P5 thus provided valuable additional evidence to the slender body of literature on the efficacy of whole word written cues as a therapy resource, e.g., Participant P.A. in Nickels and Best (1996b; the researchers' 'Study 2, Therapy 2' in Table 5.1; further elaboration provided below).

While four of the participants benefitted from the whole word orthographic cue, two individuals also significantly benefitted from initial letters (i.e., P4 and P5) on two measures each, including a count of naming accuracy per participant. This was despite the null results on naming accuracy produced by group-level analyses for the initial letter written cues, after direct comparisons were made against repeated exposure (Table 6.3). Researchers and clinicians are therefore again reminded to look beyond group averages, and whenever possible, complement group-based examination with individual analyses to avoid information loss (Nickels & Howard, 1995; Schork, 2015).

Research question (2): Duration of therapy effects for each unit of orthographic cue

Effects from whole word orthographic cues could potentially last up to four weeks, as observed in at least two participants (P4, P5). The majority of participants (all except P1) significantly improved when assessed one-week post-therapy, on at least one outcome measure (e.g., final naming accuracy). These results are in keeping with the group finding that on average, whole word written cues generated significant improvements lasting for a week and marginal benefits when assessed four weeks later.

In comparison, effects from initial letter cues (if any), potentially lasted for at least one week, as identified in the same two participants (P4, P5). However, neither participants' results differed from their baselines four weeks later (Table 6.3).³⁰

³⁰ Owing to the design of the study in which the therapies were delivered sequentially, it is inevitable that outcomes measured at four weeks or later will be after the completion of both therapies. One might therefore suggest any successful outcomes obtained four weeks later were the additive effects of the two therapies. However, if there were additive effects at play, successful sustained naming four weeks later would be even more likely obtained for items treated in the first therapy (i.e., initial letter cue) for P4 and P5 – the only two participants who showed benefits from initial letter orthographic cues after accounting for repeated exposure. Instead, it was their second therapy (i.e., whole word cue) that showed the sustained item-specific effects at four weeks later. Still to definitively rule out any additive effects, a future experiment using closely-matched samples, with each sample treated either in initial letter or whole word written cue only, is necessary.

Clinical implications based on investigations into research questions (1) and (2)

Overall, when directly comparing the individual results obtained for the two units of orthographic cues, evidence was stronger in favour of the whole word unit – It significantly benefitted more individuals (four participants for whole word versus two for initial letter written cues) and the effects were more durable (potentially maintained four weeks post-treatment for whole word versus one week for initial letter cues). In terms of the actual amount of improvement (Table 6.4), the changes in raw values yielded were respectable for both initial letter and whole word cues. Take P5, for instance, following initial letter cueing, at one-week post-treatment, her initial naming accuracy improved by 16 additional items and she was faster by 4868.4 ms. However, her improvements were even more impressive following whole word orthographic cueing: At one-week post-therapy, P5 could name an extra 17 items successfully on initial naming accuracy and 21 items successfully on final naming accuracy. Moreover, she was faster by 5517.3 ms. The latter was a further improvement of over 600 ms, when compared to that of her performance following initial letter cue therapy. In addition, no reliable difference was identified between repeated encounters and initial letter cues at four weeks post-treatment, whereas there were four participants whose improvements from whole word cues remained significantly stronger when evaluated against the corresponding results from repeated exposure (Table 6.3; also see Table 6.4 for the outcomes that remained significant against repeated exposure in raw values).

These results on the strength of the whole word written cue are consistent with the earlier group finding in Chapter 5 and results from the meta-analysis in Chapter 2. Importantly, in the context of individual case data reported elsewhere in the literature, the current findings fill in crucial gaps. The only case evidence related to whole word orthographic cueing thus far (and thereby the study whose intervention was closest to the whole word cue therapy reported here) was provided by Nickels and Best's (1996b) participant, P.A. P.A. completed the 'lexical therapy,' which involved flipping the picture card over and reading aloud the target words written at the back, should she fail to name it correctly (Table 5.1; Appendix 6.4, cells highlighted in yellow). She was able to maintain a significant naming gain of 7 items when assessed two weeks post-therapy. However, the researchers did not investigate beyond the naming accuracy dimension and two weeks after intervention. Results in this chapter therefore extend the evidence base in two ways. Firstly, there is now fresh evidence that participants improved on their speed of responses as well. At least three participants demonstrated significant RT improvements, when assessed four weeks later. The smallest amount of significant gain was registered by P4, who was faster

by 2026.4 ms, and the largest RT gain was an improvement of 3974.5 ms (P2). Secondly, naming accuracy gains can potentially be maintained for at least four weeks post-therapy. Based on final naming accuracy, this could be as high as a significant gain of 12 additional items compared to baseline (P5). These are encouraging values in favour of whole word orthographic cueing, lacking in previously published works.

Thus, when choosing between which orthographic unit to employ for cueing therapy, in the absence of outstanding reasons (e.g., client's explicit preference for initial letters), clinicians could consider applying the whole word cue.

Secondary research question (3) and associated clinical implications: Relationship between each participant's baseline performance on particular cognitive skill(s) and the ability to benefit from each unit of orthographic cue

Part of patient-centred care involves applying the appropriate therapy according to the profile of each client. To explore this, for every cue type, each individual's findings (i.e., which unit of cue was effective) was linked to his/her language and cognitive performance on standardised assessments gathered at baseline. Based on the therapy cue outcomes, three groups of participants could be identified according to the therapy cues that worked: (1) Participants who did not benefit from either orthographic cue (i.e., P1), (2) Participants who only benefitted from the whole word orthographic cues (i.e., P2 and P3), and (3) those who benefitted from both initial letter and whole word orthographic cues (P4 and P5).

By arranging the individuals in ranked order according to their baseline performances alongside their therapy cue groupings (Figure 6.5), two observations could be made. It should however be emphasised that this examination into secondary research question (3) was exploratory, and its associated findings would require corroboration by future research using the regression approach with larger samples.

(1) In order for orthographic cues to deliver therapeutic effects, participants should at least be able to process written word forms, i.e., they are not severely impaired across all three orthographic processing paths 'A' to 'C' in Figure 6.1. This makes intuitive sense, as an individual with premorbid dyslexia might not benefit much from written cues either. Case evidence for this is supplied by P1. From a theoretical perspective, all three possible orthographic pathways (paths A, B, and C in Figure 6.1) were impaired for P1. On the basis of P1's data in relation to others', Figure 6.5 highlighted a similar disparity in language profile with P1 being the lowest scorer on any language tasks exclusively reliant on visual word processing. Similar patterns based on his performance on the remaining tasks could not be identified: For instance, P1's was equally weak on 'CAT 4: Recognition Memory' as P2 (for

whom whole word cue was effective) and P4 (for whom both initial letter and whole word cues were effective), or that he was superior to others on the 'WMS: Digits Backward Recall'. Combined, these suggested that it was perhaps the ability to process written words (including the ability to read them aloud), rather general cognitive skills, which was crucial for orthographic cues to operate.

(2) To benefit from initial letter written cues, relatively good visual short-term memory measured by 'Corsi Block Tapping Test' could be important. It was the only task that differentiated between participants who found them effective (P4 and P5) against the rest who did not (Figure 6.5). A theoretical account for this finding is not immediately obvious. Perhaps initial letters being visual forms are more sensitive to visual-based cognitive skills, and 'Corsi Block Tapping Test' captures the ability to hold and retrieve visual information, possibly necessary for visual cues to be processed adequately. Exact details on how visual short-term memory contributes towards the therapeutic process of initial letter written cueing, beyond a temporary visual storage of the grapheme, is unclear at this stage. Ktori, Grainger, and Defau (2012) provided data from healthy adult readers, on a specialised encoding mechanism within visual short-term memory to decode letter strings; this visual short-term memory mechanism could be impaired to varying degrees in people with aphasia post-stroke, contributing to differences in benefits gleaned from initial letter cues.

Generalisation to unseen controls

For three out of the five participants, there were no significant changes to the performance on unseen controls, suggesting that perhaps, for most individuals with anomia, the effects from orthographic cueing are item-specific. However, results from two participants (P4 and P5), suggested limited generalisation could occur. Specifically, P4 showed marginally significant improvements naming unseen controls on final naming accuracy and RT, when he was assessed on Week 13 (seven weeks after the end of his first therapy, i.e., written initial letter cueing; and four weeks after the second therapy, i.e., written whole word cueing); whilst P5 significantly improved in her RT taken to name unseen controls when assessed on Week 7 (one week after the end of her first therapy, i.e., written initial letter cueing; see Figure 5.1 for timeline and Table 6.3 for results). This is in the context that no significant changes were detected in these two participants on the nonword repetition control task (Section 6.3.2.5), suggesting that the limited generalisation observed could not be attributed to general improvement across time. The two participants happened to be the two who benefitted from both initial letter and whole word cues and showed better performance on 'Corsi Block Tapping Test' (see secondary research question 3). Thus,

good visual short-term memory could potentially play a role. It is unclear how visual short-term memory contributed to generalisation, or whether this associative finding was coincidental. The implications of limited generalisation (if any) from orthographic cueing will be covered in the General Discussion (Section 8.7, p. 233).

Incidental evidence on repeated exposure without therapy

Finally, the results suggested that the participants were capable of improving through repeated attempts at naming, in the absence of cues. Repeated exposure conditions were designed as controls in this study, and evidence on their effectiveness was thus incidental. These naming improvements were consistent with previous literature (e.g., JAW in Nickels, 2002a). However, not all participants are influenced by the effects of repeated naming encounters (e.g., P3). This is also in keeping with published research (see Creet et al., 2019, in which only four out of 23 participants improved naming based on repeated exposure). It is still unclear what enables some participants to improve following repeated exposure while others do not, though it has been hypothesised that the greater the number of attempts at naming, the higher the likelihood of improving on that item, when individuals detect that their specific naming attempts are correct. According to Creet et al. (p. 1220), their learning system would then be switched on, and the stimulus (picture) - response (spoken word) pairing would be correspondingly strengthened. This process assumes good ability to monitor response accuracy and adjust responses accordingly, which are skills linked to the performance on executive function measured by the Wisconsin Card Sorting Test (WCST; e.g., Fillingham et al., 2005). This hypothesised mechanism is however not supported by the data presented here. While WCST was not administered, the Brixton Spatial Anticipation Test was included, and both tests measured the same latent construct related to the flexible detection of rule changes. Participants who scored better on either of the two tests, should be in a stronger position to improve following repeated exposure. Based on Figure 6.5, P3 who failed to benefit from repeated exposure, had the highest score on the Brixton Spatial Anticipation Test, contradicting Creet and colleagues' (2019) hypothesis. Nevertheless, people with aphasia form a very heterogeneous population, and P3 could be an outlier. Exactly why repeated exposure is beneficial for some and not others, will require further investigation using a larger sample with diverse cognitive skills. An attempt to identify locus of repeated exposure in the cognitive neuropsychological model will continue in Chapter 7 through the lens of error analyses.

In conclusion, individual-level analyses clarified the group-level evidence for orthographic prompts. In particular, they provided corroborating data on the efficacy of

whole word written cues – that four out of the five participants significantly benefitted from it suggests this cue format holds promise.

Chapter 7

Evaluating the changes in error types for initial letter and whole word orthographic cueing – Individual participant error analyses

7.1 Introduction

This chapter builds on the individual accuracy and RT outcomes obtained in Chapters 5 and 6 by tracking the changes in error types before and after therapy. Using error analyses as secondary analyses to increase understanding, e.g., to further evaluate a therapy's efficacy, has gained traction over the years (Kendell et al., 2013).

To clarify, error analyses here do not refer to diagnostic evaluation of naming errors to infer what could be the locus/loci of deficit in a person with anomia at a specific time point (Fromkin, 1971; Fridriksson, Baker & Moser, 2009). Instead, error analyses in this chapter refer to the comparison of error rates across time, particularly before and after therapy. Table 7.1 summarises word-finding therapy research that have utilised this form of error analyses to explore whether a treatment had any impact on specific error types. In table 7.1, only therapies involving language forms as input and whose research participants are adults with post-stroke anomia are included. These boundaries are in keeping with the literature reviews in earlier chapters. Hence, treatments involving non-language forms, e.g., having participants to draw pictures for self-cueing (one of the experimental conditions in Davies, Farias and Baynes, 2005), or research involving other participant groups, like people suffering from word-finding difficulties due to depression (Georgieff et al., 1998), are excluded.

Based on table 7.1, a recurrent finding emerges: Significant decreases of 'no response(s)' are commonly observed when error types are compared pre- and post-therapy phases across different treatments. Outcomes for semantic, phonological or other error types, on the other hand, were generally not significant. This significant decrease in 'no response(s)', for instance, is detected following the phonomotor treatment (Minkina et al., 2016, 2019), as well as after modified semantic features analysis (Hashimoto, 2012). Interestingly, none of these pieces of research examined whether the decline in 'no responses' could be a consequence of mere repeated encounters with the same items, i.e., priming, rather than actual changes induced by the therapies (see JAW in Nickels, 2002a, mentioned in Chapter 3). To properly distinguish, in this chapter, outcomes on 'no response(s)' for the cue conditions and their attendant control conditions of repeated exposure will be jointly analysed. This should clarify whether it is indeed the cueing therapy

Table 7.1. Summary of spoken word-finding facilitation/therapy studies that included error analyses when evaluating a therapy's efficacy. All studies focus on spoken picture naming accuracy as the primary outcome and involve only individuals with aphasia post-stroke.

Study (Chronological)	Description of facilitation/therapy and error types analysed	Main Findings (comparing post-facilitation/therapy vs. final baseline, except for Stimley and Noll, 1991) based on picture naming scores
Stimley and Noll (1991) <i>n</i> = 20 adults with aphasia (9 males, 11 females; age 30 to 83 years; 1 to 156 months post-stroke)	Facilitation technique: Pre-stimulation with either semantic or phonological cues (e.g., 'This is something you can ring,' for a 'bell'; 'This is something that starts with [insert phonemic cue]') Treated language: English Treatment frequency: One session Error types analysed: Semantic, phonological, unrelated real word, perseverations Although other error types were coded, e.g., circumlocutions, part-whole (i.e., named a part of the picture instead of the whole picture), these additional errors were not analysed.	• Significant increase in treated items correctly named post-facilitation with either semantic or phonological cues vs. neutral condition across participants. Error analyses (mean proportion of errors) conducted for treated items across participants: • Significant increase in semantic errors following the semantic cue compared to either the phonological cue or neutral condition. • Significant decrease in unrelated word errors following the semantic cue compared to either the phonological cue or neutral condition. • Significant increase in phonological errors following the phonological cue compared to either the semantic cue or neutral condition. Key implication(s) inferred from error analyses: • Phonological word form access facilitated by top-down processes in semantic cueing (thus less 'unrelated' word errors) and by bottom-up processes in phonological cueing (failure to inhibit inappropriate but adequately activated phonemic representations, thus increase in phonological errors).
Davis, Farias and Baynes (2005): <i>n</i> = 8 adults with aphasia (6 males, 2 females; age 50 to 71 years; 1 to 13 years post-stroke)	Facilitation technique: Orthographic self-cueing by attempting to write out the target word Treated language: English Treatment frequency: One session Error types analysed: Semantic, semantic-related description, phonological, unrelated real word, perseverations, nonword (authors used the label 'neologism'), no response(s)	• No statistical tests conducted on the naming accuracy outcomes of items. Error analyses (number of errors) conducted for treated items across participants: • No significant differences found in any of the errors when assessed immediately following the participants' attempt to self-cue via writing, compared to baseline naming. Key implication(s) inferred from error analyses: • None inferred in the paper.
Gordon (2009): Therapy 1 and Therapy 2 <i>n</i> = 2 adults with aphasia (Ag1: 18-year-old female,	Therapy 1 technique*: Semantic cue(s) generation – Modified semantic features analysis; generate three semantic features , i.e., category, appearance, and 'what it feels like'. Therapy 2 technique*: Context treatment – Listen to a story related to target word. Retell the story afterwards, with cues	• *Because both therapies were conducted concurrently (modified semantic features analysis and contextual therapy), it was not possible to tease apart the effects of each therapy. Joint outcomes thus reported below. • Only visual analyses based on the eyeballing of descriptive statistics (reported in percentages) done. No statistical tests conducted.

14 months post-stroke; Ag2: 44.5-year-old male, 9 years 2 months post- stroke)	from therapist [no details given on the types of cues used or when to apply]. Treated language: English Treatment frequency: Duration per session and frequency of sessions per week not stated. - Ag1: 17 sessions (including probes) over 12.5 weeks - Ag2: 15 sessions (including probes) over 8 weeks Error types analysed: Semantic, phonological, mixed semantic and phonological, unrelated, miscellaneous (labelled as 'Other')	Improvement on treated and untreated items for each participant, based on visual eyeballing. Error analyses (percentage of total errors) combined error types across the two therapies (including therapy 2 below) but conducted at the participant-level: - Ag1: Reduction observed in phonological, mixed semantic and phonological, and unrelated error types, based on visual eyeballing. - Ag2: Greatest reduction observed in 'Other' responses (i.e., circumlocutions and no responses), based on visual eyeballing. Key implication(s) inferred from error analyses: '... weren't enough instances of each error type to permit a reliable analysis' (p. 654).
Hashimoto (2012): Therapy 1 and Therapy 2 <i>n</i> = 2 adults with aphasia (FP: 66-year-old female, 60 months post-stroke; KJ: 33-year-old female, 18 months post-stroke)	Therapy technique 1: Semantic cue(s) generation – Modified semantic features analysis; generate three semantic features related to the target word, i.e., group, physical-sensory properties, association Therapy technique 2: Phonological cue(s) generation – Modified phonological component analysis (Leonard et al., 2008); generate three phonological features related to the target word, i.e., number of syllables, initial phoneme, rhyme Treated language: English Treatment frequency: Two sessions (45-60 mins each) per day, twice per week. Termination criterion: naming accuracy ≥ 80% over three consecutive sessions. - FP: 29 (Set 1) and 14 (Set 2) sessions - KJ: 11 sessions respectively for Sets 1 and 2 Error types analysed: Semantic, unrelated real word, no response(s)	Effect sizes (Busk & Serlin, 1992) were computed only for treated items. No further statistical tests conducted for treated or untreated items. Therapy 1 (Set 1 = immediately post-therapy, Set 2 = six weeks post-therapy) FP: Large (Set 1) and medium (Set 2) effect sizes KJ: Small (Set 1) and medium (Set 2) effect sizes Therapy 2 (Set 1 = immediately post-therapy, Set 2 = six weeks post-therapy) FP: Small (Set 1) and small (Set 2) effect sizes KJ: Large (Set 1) and small (Set 2) effect sizes Error analyses conducted for treated items at the participant level: - Number of errors for each error type for first five treatment sessions were compared to the number of errors in the last five sessions. FP, with same pattern of results found in therapies 1 and 2 respectively: Sets 1 and 2: Significantly fewer 'no responses' in the last five sessions. Set 2: Significantly more 'semantic errors' in the last five sessions. KJ, with same pattern of results found in for therapies 1 and 2 respectively: Set 1: Significantly fewer 'no responses' in the last five sessions. Set 2: No difference was found in each of the three error types on post hocs. Key implication(s) inferred from error analyses: - Evolution of errors suggest increased lexical access post-therapy in both participants, since the 'presence of no responses represents failures of access of the targeted lexical representation' (p. 540).
Bose (2013)	Therapy technique: Phonological cue(s) generation – Phonological component analysis (Leonard et al., 2008); asked to generate five phonological features related to the target	Only visual analyses (≥ two consecutive post-therapy data points falling outside either ±2 SDs of the baseline data) done. No statistical tests conducted on the naming accuracy outcomes of treated items.

n = 1 adult with aphasia (FF: 77-year-old male; 4 years post-stroke)

word, i.e., the rhyme, initial phoneme, first sound associate (i.e., another word beginning with the initial phoneme), final phoneme and number of syllables

Treated language: English

Treatment frequency: 32 sessions (45-50 mins each) held three times per week

Error types analysed: Semantic, phonological, mixed semantic and phonological, nonword (author used the label 'neologism'), miscellaneous

- **Improvement on treated items post-therapy** (based on + 2 *SDs*) at maintenance and follow-up phases. **Exact time length between testing and the last therapy session for each list is unclear** for maintenance phase.
- **No improvement** (based on + 2 *SDs*) on **untreated items post-therapy**. **Time length between post-therapy testing and the baseline was not provided**.
- Error analyses (proportion of errors) were *not conducted* for treated items.
- **Error analyses** (proportion of errors) for **untreated items** (referred to as Philadelphia Naming Test by Bose as she used it to measure generalization): **Significant decrease** in **neologisms** produced post-therapy. **Significant increase** in **miscellaneous errors** (real words) post-therapy. Changes in the other errors were not significant. **Time length between post-therapy testing and the baseline was not provided** though.

Key implication(s) inferred from error analyses:

- Simultaneous decrease in neologisms and increase in real words (not always correct) responses suggest an '**improved access to phonological forms**' post-therapy (p. 593). But note error analyses were not conducted for treated items.

Kendell, Pompon, Brookshire, Minkina, and Bislick (2013)

n = 10 adults with aphasia (gender breakdown not provided; age 27 to 72 years; 14 to 211 months post-stroke)

Therapy technique: Phonomotor treatment (see pp. S243-S244, Kendall et al., 2013) – **Multimodal therapy** in two stages.

- **Stage 1:** explore **sounds in isolation**. E.g., shown mouth picture of sound, imitate sounds after therapist.
- **Stage 2:** explore **sounds in combination**. E.g., produce sounds according to the series of mouth pictures or letters.

Treated language: English

Treatment frequency: 10 sessions (60 mins each), held twice per day, five days per week.

Error types analysed: Semantic, phonological, mixed semantic and phonological, unrelated real word, nonword, omission (authors defined as 'no response, with semantic description' on p. S242).

- **Significant increases** in the naming of **treated items** found **one-week and three months post-therapy across participants**.
- **No significant changes** found for the naming of **untreated items**.

Error analyses (mean proportion) on treated and untreated items across participants:

- **Treated items:** Near-significant **decrease** in '**omission**' errors produced **one-week post-therapy** ($\alpha = .01$ after correcting for Bonferroni)³¹.
- **Untreated items:** Near-significant **increase** in '**mixed semantic and phonological**' errors produced **one-week post-therapy**.

Key implication(s) inferred from error analyses:

- 'Reduction of omission errors following treatment for trained items is consistent with movement from a more severe naming deficit to a milder form' (p. S246).
- Mentioned improved activation of semantic processing (p. S246); basis of inference is unclear, as changes in semantic errors were not significant ($ps > .15$; p. S244).

Minkina et al. (2016)

Therapy technique: Phonomotor treatment – Multimodal therapy in two stages. See above for its description.

- **Significant increases** in the respective naming of **treated and untreated items**; both were found **one-week and three months post-therapy across participants**.

³¹ Although Kendall et al. wrote that 'a decrease in omission errors and an increase in semantic errors for trained items were observed three months post-treatment' (p. S246), neither claim is consistent with their results (pp. S 244 - S246). In the paper, they identified a near-significant change in omission errors in *one-week* post-therapy, not three months. However, they did not identify any significant outcomes for semantic errors when the mean error proportions were statistically compared before and after therapy.

n = 24 adults with aphasia (10 of them whose data were analysed in Kendell et al., 2013; 14 males, 10 females; age 26 to 78; 8 to 211 months post-stroke)

Treated language: English

Treatment frequency: 10 sessions (60 mins each), held twice per day, five days per week.

Error types analysed: Semantic, phonological, omission (definition provided above). Phonological errors are defined as those sharing at least two sounds with the target word (p. 967).

Although other error types were coded, i.e., mixed semantic and phonological, unrelated real word (authors used the label 'semantic-unrelated'), and nonword, these three error types were not analysed due to 'their low occurrence' (p. 969).

Minkina et al. (2019)

(*n* = 28; 15 males, 13 females; age 40 to 90; 10 to 298 months post-stroke)

Therapy technique: Phonomotor treatment – Multimodal therapy in two stages. See above for its description.

Treated language: English

Treatment frequency: Estimated 10 sessions (45-50 mins each, with 10-min break in between; 56-60 hours of treatment in total), held twice per day, over six to seven weeks

Error types analysed (Table 3 in Minkina et al., 2019, p. 4087):

- **Target-related:** Semantic, phonologically-related real word, mixed semantically and phonologically-related real word, phonologically-related nonword
- **Target-unrelated:** Description, unrelated real word, phonologically-unrelated nonword, miscellaneous
- **Omission:** Fragment (i.e., minimally CV or VC response) and no response

Subgroup error analyses (mean proportion of errors) conducted for **treated and untreated items** across (a) participants with low WAB scores (high severity), (b) participants with low BNT scores (poorer lexical retrieval):

- **Significant decrease** in 'omission' errors produced **three months post-therapy** in **each subgroup analysis** (Table 8 in Minkina et al., 2016).
- **Results for other error types** from parallel subgroup error analyses across those with (a) high WAB scores, (b) high BNT were respectively **not significant**.

Key implication(s) inferred from error analyses:

- '... **decrease in proportion of omission errors** may be due to increased semantic activation in the network as a result of feedback from strengthened phonological representations ... suggesting our **treatment strengthened the linguistic system holistically**.' (p. 976)

- **Significant increase** in the naming of **treated items** when assessed **immediately and three months post-therapy** across participants.
- **Significant difference** naming **untreated-linguistically related (Untreated-LR) items** when assessed **three months post-therapy** across participants.
- **No significant difference** naming **untreated-linguistically unrelated (Untreated-LU) items**, be it immediately or three months post-therapy across participants.

Error analyses (mean proportion of errors) conducted for **treated and untreated items respectively** across participants:

- **No significant differences** found in **all target-related error types** for treated, untreated-LR or untreated-LU items, be it immediately or three months post-therapy.
- **Significant decrease** in **target-unrelated description errors** for **treated items** when assessed **immediately post-therapy**.
- **Significant decrease** in **omission errors** for **treated items** when assessed **immediately post-therapy**.
- **Rest of the results for error types** were **not significant**.

Exact time length between testing and the last therapy for 'immediately' is unclear.

Key implication(s) inferred from error analyses:

- When data for target-unrelated description errors were recoded, it appeared result was driven by single verbs, adjectives etc. *semantically related* to the target word.
- **Not possible to infer what caused omission errors** (p. 4094). Possibly, **therapy impacted the entire word retrieval system**. Thus, **with participants making responses**, there is a **decrease in omission errors**, which include 'no responses.'

implemented in my study that resulted in this error change (if any) or is it merely a function of repeated priming that led to a fall in 'no response(s)'.

One reason for conducting error analyses is because they may uncover patterns that are not identifiable through evaluations of accuracy data. In Table 7.1, this discrepancy is reported in Bose (2013): Although no significant results were found comparing the pre- versus post-therapy accuracy outcomes for the untreated items ($p = .81$; p. 590), there was a significant decrease in the proportion of neologisms in the same period for the same set of items. Bose thus concluded that 'the quality and quantity of errors (in addition to accuracy of response) may provide a sensitive outcome measure to determine therapy effectiveness' (p. 593). Bose's discrepant finding is an anomaly. As seen in Table 7.1, similar drastic contrasts in findings between error analyses and accuracy outcomes (significant in one, null in the other) are uncommon. It is more likely for studies to obtain significant accuracy outcomes, which are then mirrored by significant changes in specific error types for the same set of items (e.g., Minkina et al., 2019). Curiously, in Bose (2013), error analyses were also selectively applied only to untreated items. Bose's conclusion might carry more weight had she conducted error analyses for *both treated and untreated* items, and that if contrasting outcomes in accuracy and error outcomes were unveiled for the treated items as well.

Finally, two caveats come to fore when discussing error analyses: (1) The first is that it is not always possible to directly link a change in error type to a specific mechanism. For instance, an increase in semantic errors can be attributed to either an incorrect semantic neighbour being selected over the target due to the 'noise' at the semantic features level, or because of a post-semantic deficit whereby there was a failure to access the target phonological representations (Caramazza & Hillis, 1990). In the latter, the semantically-related phonological representation with the strongest activation will be produced instead. Both scenarios are plausible when participants demonstrate both semantic and post-semantic deficits in their initial assessments (e.g., P3 in this intervention study; see section 6.2.2.3). Another example of multiple causes will be the 'no response' errors. They could arise as follow: Firstly, there could be a complete failure to activate the semantic nodes, or in the case that semantic nodes are successfully activated, there might be a faulty link between the semantic node and the corresponding lexical entry. A third possibility is that the null response is simply because of good self-monitoring, i.e., the individual inhibits spoken responses to avert overt erroneous answers.

(2) A second caveat pertains to the small number of errors typically made by an individual for each error type in most studies. This may lead to inconclusive results due to

floor effect. To quote Gordon (2009) whose error analyses yielded null findings, ‘... [there] weren’t enough instances of each error type to permit a reliable analysis’ (p. 654). The reasons for the small numbers are likely logistical, e.g., constrained by the limited number of naming stimuli to begin, or simply because the participant rarely makes that category of error. The limited number of stimuli is sometimes a necessary constraint given ethical concerns – It is not ethical to subject a fatigued individual or a person with attention deficits to a large number of items, for the sake of creating sufficient error numbers. Lethargy and reduced attention are commonly experienced by individuals after a stroke (Paciaroni & Acciarresi, 2019). Having a large number of stimuli may dissuade the exhausted participant, risking them pulling out or delivering suboptimal performances. All these render results unreliable. In this therapy study, for each condition, I have included as many items as humanely possible for each participant to tolerate with breaks to try address this caveat.

These limitations notwithstanding, there is still a chance that error analyses could add depth to the existing findings reported in the earlier chapters, as documented by the untreated items in Bose (2013). Based on literature (Table 7.1), the following were thus hypothesised. Where specific error types are mentioned, they will be defined/elaborated in Table 7.2 that follows.

Hypothesis (1): The rate of making *target-related errors* after therapy follows an *increasing* trend; this trend should be observed for items treated in the cue conditions only, but not for the control conditions (e.g., repeated exposure condition).

Target-related error types are errors that are associated with the target word in a discernible manner. The association can be semantic, phonological, orthographical, a mixture of semantic and phonological, or a direct indication that a spoken word is not the target word (Table 7.2). For the latter, stating a response is invalid implies a cognizance that the word produced is not related to the correct answer. The increase in target-related errors suggests the treatment is showing signs of efficacy because that individual, rather than making a random error, has produced an erroneous response that shows some linguistic resemblance to the correct word. Examples of research identifying increase in target-related errors following treatment include Stimley and Noll (1991), Hashimoto (2012) and Kendall et al. (2013).

Hypothesis (2): The rate of making *target-unrelated errors* after therapy follows a *decreasing* trend; this trend should be observed for items treated in the cue conditions only, but not for the control conditions.

Target-unrelated error types are errors that are not associated with the target word in any discernible ways. These include real or nonsense words that do not share an identifiable relationship with the target response (Table 7.2). If a treatment is efficacious, the individual should not make random-like target-unrelated responses, i.e., responses that do not resemble to the target word in any way. Examples of research that document a decline in target-unrelated errors following treatment are Stimley and Noll (1991), Hashimoto (2012), and Minkina et al. (2016, 2019).

Hypothesis (3): The rate of making *null-type errors* (e.g., no response) after therapy follows a *decreasing* trend; this trend should be observed for items treated in the cue conditions only, but not for the control conditions.

In the absence of an overt response, it is not possible to classify whether the answer is related to the target word or not. Nonetheless, it is hypothesised that should a particular cueing treatment be effective, the ability to provide a naming response instead of providing none, should be improved, following that cue therapy. Examples of research supporting a decline in null error types after therapy include Hashimoto (2012) and Minkina et al. (2016, 2019; whereby the authors classify ‘no response(s)’ under ‘omission’ errors).

7.2 Method

7.2.1 Participants and data collection

Data from the intervention study (Chapters 5 and 6; $n = 5$ adults with chronic anomia) were coded for error types and analysed. Following previous studies, e.g., Hameau, Biedermann, Fieder and Nickels (2020), only initial naming responses were analysed for their errors. Details on the participants and data collection procedure can be found in sections 5.2.2 and 5.2.3. To recap, all participants named picture in these five conditions:

- Condition (1): Initial Letter Cue: Orthography Only
- Condition (2): Repeated Exposure: Initial Letter Cue (i.e., Untreated items encountered during ‘Initial Letter’ orthographic cue therapy)
- Condition (3): Whole Word Cue: Orthography Only, and
- Condition (4): Repeated Exposure: Whole Word Cue (i.e., Untreated items encountered during ‘Whole Word’ orthography cue therapy)
- Condition (5): Unseen Control (i.e., Untreated items not encountered during either therapy)

Conditions (2), (4) and (5) are control conditions. As mentioned in this chapter's introduction, conditions (2) and (4) were included to distinguish cue therapy-specific effects and outcomes that could simply arise from repeated encounters alone (see section 5.2.3.2 for details on these conditions).

7.2.2 Classifying naming errors using an error coding scheme

Firstly, all naming errors produced as the first naming response by each individual at each time point were coded at the item level. An error coding scheme adapted from Best, Nickels, and Williamson (2005) was used because it was developed to classify picture naming errors found in people with aphasia; also, it is comprehensive – errors like 'orthographic error' and 'no (specific) response' are accommodated. Coding for orthographic errors is meaningful in this thesis even if numbers turn out to be small, as variants of the written cues are manipulated as experimental conditions, so it makes sense to check if there are direct fallouts in terms of error type. Interestingly, orthographic errors tend to be omitted in aphasia naming error coding schemes, e.g., Kohn and Goodglass (1985), Stimley and Noll (1991), or Philadelphia Naming Test (Moss Rehabilitation Research Institute, 1996; Roach, Schwartz, Martin, Grewal, & Brecher, 1996), although there is evidence of orthographic errors in naming by adults with aphasia (e.g., Dean, 2010). Finally, Best, Nickels, and Williamson's coding scheme has been fruitfully applied in other aphasia naming research, though not necessarily stroke-related, like van Scherpenberg, Fieder, Savage, and Nickels (2018) and Hameau et al. (2020), thus increasing confidence.

Nonetheless, some adjustments to Best et al. (2005) are necessary. (1) Their original aim was to develop a detailed system to classify semantic errors, which explains why they created nine semantic error categories for single-word errors. This fine-grained cataloguing of semantic errors is applicable for research delving into semantic deficits, e.g., participants with primary progressive aphasia, whose naming responses are predominantly semantic errors (van Scherpenberg et al., 2018), but it does not apply to this thesis. Separating into many semantic categories here is disadvantageous because the final count of each semantic error subtype is likely very small, which could in turn mask results/patterns. To avoid Type-II errors for potential semantic-related findings, the nine subtypes are collectively coded as 'semantic errors.'

(2) Best et al. (2005) coded for 'visual error' (e.g., 'cookie' for a picture of 'potato') and 'reading orthography in picture.' These errors were collectively coded as 'miscellaneous errors,' together with errors that simply could not be categorised. This is

because these two errors are rare (if any): (1) There are no written words in any of the pictures for 'reading orthography in picture' to manifest. (2) There were very few visual errors made by the five participants across the six assessment time points³².

(3) In Best et al. (2005), only single word responses were coded. In my sample, sometimes participants produced multiword utterances during picture naming. To accommodate, 'Semantic: Circumlocution, description' and 'Other errors: Multiple words; no link to target' were included to represent multiword responses that are semantically related and unrelated to the target word respectively.

(4) Lastly, for purposes of clarity and to be consistent with the literature, e.g., Minkina et al. (2019), and hypotheses, I have grouped each error type into three broad categories, depending on whether it is target-related/unrelated or whether it is a null-type error. There are thus seven target-related errors, four target-unrelated ones and one null-type error. This categorisation will help identify if a specific cue condition increases proximity towards the target word or not, with errors related to the target word indicative of closer proximity.

The 'naming error types' coded for are listed in Table 7.2.

7.2.3 Fidelity check

Similar to the facilitation and therapy studies (sections 3.2.4.5 and 5.2.3.4), 20% of the naming error data were independently verified by a licensed speech and language therapist fluent in English and external to the research team. The same rater in Chapter 5 is engaged here. Likewise, the data were randomized, rendering the rater blind to the specific condition as well as the time point from which the naming response arose. Any differences between raters when classifying error types were resolved by discussion. Overall, the interrater agreement for categorising error types was high at each time point, achieving at least 90% agreement per assessment session [Average percentage agreement across sessions = 93.84%]. The interrater Krippendorff's α -values were similarly respectable, with each session surpassing .81 [Average α -value across sessions = .899](Krippendorff, 2004; Freelon, 2010, 2013).

³² The highest absolute number of visual errors is made by any participants at a particular time point for a condition is P4, where two visual errors were identified in the 'Whole Word Cue' condition, such that '2/23' initial naming responses and '2/27' final naming responses were identified at 'Week 13' and 'Week 2: Baseline 2' respectively (e.g., 'ice' when naming 'ashtray'). For all other instances (across participants, conditions, and time points), the maximum absolute number of visual errors is '1' (if any). P2, in fact, did not make any visual errors throughout the study.

Table 7.2. The three categories and their specific types of object-naming errors in aphasia, adapted from Best, Nickels, and Williamson (2005).

Naming error type	Definition / Example
a. Target-related error	
Semantic error	A spoken single noun response that is related to the target word in terms of meaning. The shared meaning can take the following forms: • Category coordinate (e.g., 'orange' for 'apple') • Superordinate (e.g., 'plant' for 'tree') • Subordinate (e.g., 'shirt' for 'clothes') • Associated (e.g., 'medicine' for 'doctor')
Semantic error: Circumlocution, description	A multiword spoken explanation or a recount of the item's appearance or an experience with it.
Phonological error: < 50%	A spoken single-word response that shares less than 50% of the phonemes in the target word. A typical example is a naming response made up of the first phoneme/syllable of the target word. This includes what is coded as "fragment" in the Philadelphia Naming Test (Moss Rehabilitation Research Institute, 1996). E.g., /t/ for 'tree' (/tri:/)
Phonological error: ≥ 50%	A spoken single-word response that shares at least 50% of the of the phonemes in the target word. E.g., /kʌm.pə/ for 'compass' (/kʌm.pəs/)
Semantic/Phonological error	A spoken single-word response that is semantically-related to the target word, whilst sharing at least 50% of the target phonemes. E.g., 'lemon' for 'melon'
Orthographic error: < 50% ³³	A written single-word response with less than 50% of the target word spelt out. E.g., 'g' for 'garlic.' Includes finger-spelling.
Rejection: Reject incorrect response(s)	Declining a spoken response to be the target word, when the response is indeed inaccurate. The declining (e.g., no, sorry, can't say it) should take place within two seconds from the end of the inaccurate spoken naming attempt. E.g., 'cat, no' or 'cat, can't say it' for 'dog' or saying 'cat,' while shaking head for 'dog.'
b. Target-unrelated error	
Other errors: Real word; no link to target	A spoken response that is a real word in the lexicon, but not related to the target word. E.g., 'duck' for 'compass'
Other errors: Nonword; no link to target	A spoken response that is not a real word in the lexicon and is not related to the target word. E.g., /vu/ for 'compass'
Other errors: Multiple words; no link to target	A multiword spoken response that is not related to the target word.
Miscellaneous error	This includes any errors that cannot be categorised into any error types listed here.
c. Null-type error	
No response	Nil response, or simply saying 'no,' or just a quiet gesture indicating 'no,' e.g., shaking one's head. These are categorised as 'Omission' in Minkina et al. (2019).

³³ Note there is only one response in the entire naming data that is an orthographic error. It was produced by P4 and her response was less than 50% of the target spelling. There are no orthographic errors that would could be classified as "Orthographic error: ≥ 50%" in this sample.

7.2.4 Analysing with the non-parametric weighted trend test

Analyses were conducted at the individual level for each error type per condition. A non-parametric weighted trend test for Poisson distributed rates was used to examine the changes in error rates across the six time points. As a refresher, the six time points refer to assessments conducted at the three baselines, one-week after therapy 1 (Post-therapy assessment 1 on Week 7), one-week after therapy 2 (Post-therapy assessment 2, Week 10) and four-weeks after the completion of both therapies (Post-therapy assessment 3, Week 13). Rates were used because the number of possible errors varied across time points (i.e., the denominators were different at every time point). The Poisson distribution is suitable because the data were collected within a circumscribed time interval of 13 weeks; also, it is fair to claim that the making of the errors are discrete events (i.e., each error type is produced one at a time, not simultaneously with another error type) and each error event is independent (Boswell, 1966).

All baseline data were assigned the same weights, as were all post-therapy data. Thus, in the case of target-related errors, like 'Rejection: Reject incorrect response(s),' because correct rejections of inaccurate nouns are likely to increase after therapy: For therapy 1, each pre-therapy data point would be weighted '-1' and each post-therapy data point, '+1'. For therapy 2, which has four baselines and two post-therapy sessions, each pre-therapy data point would be weighted '-1' and each post-therapy data point, '+2'. Reverse coding was applied for target-unrelated errors since their proportion should decrease after therapy, e.g., for therapy 1, each pre-therapy data point would be weighted '+1' and each post-therapy data point, '-1'. Poisson trend test has been successfully applied in other aphasia research, involving behavioral count data employing similar rationales for weight assignment (Beeke et al., 2015; Best et al., 2016).

When significant outcomes were identified across the six time points, post hoc analyses were conducted, first by comparing the rates across the baselines, followed by the comparing of rates at each post-therapy time point against that of the final baseline. For target-related errors in the case of therapy 1, when baselines were compared, the data points were weighted as '-1, 0, +1,' whilst for therapy 2, the weights were '-1, -1, +1, +1.' When directly comparing the final baseline against each post-therapy time point, the weights were '-1, +1' respectively. Reverse coding was applied for the weights for target-unrelated errors, e.g., baseline weights were '+1, 0, -1' for therapy 1 and '+1, +1, -1, -1' for therapy 2. Bonferroni's corrections were applied to all post hoc tests.

7.3 Results

Tables 7.3 and 7.4 summarise the results obtained respectively for target-related, target-unrelated and null-type errors made by each participant. The complete set of results is published in Appendix 7. The key findings are highlighted below. Given the nature of error analyses, i.e., the count frequency of specific error types can be small, there is a chance results are spurious. Take for instance, P5's performance in the 'Whole Word Cue' condition (Table 7.3): Although P5's change in 'Phonological error: < 50% similar [to target]' under the 'Whole word cue' is statistically significant [$p = .034$], the raw values however reveal that this significant outcome was generated by the emergence of *just one* error at the final time point, while raw values were zeroes for the other time points. It is thus possible that this 'significant result' was a function of floor effects, since a solitary '1' could spur a significant outcome. To avoid related spurious findings, Howard's advice (David Howard, personal communication) of being mindful if the raw numbers are less than '5' is adhered. For this reason, only significant results whose raw values are ≥ 5 on at least three of out the six time points (i.e., present in half of the number of time points) are interpreted. In Tables 7.3 and 7.4, as well as the full results in Appendix 7, the interpretable significant outcomes are lettered in 'green' for *increasing* trends and 'red' for *decreasing* trends.

7.3.1 Target-related naming errors

Only one piece of result approached significance and was interpretable: An increasing trend was observed for 'Semantic' errors for the 'Unseen Control' condition in P1 [$p = .052$]. Post hoc testing showed that the increase in semantic error rate made seven weeks post-therapy was marginally significant, compared to that of the final baseline, after adjusting for Bonferroni's [$p = .046$]. The rest of the results were not significant (Table 7.3, Appendix 7). A baseline check on the rates of making 'Semantic' errors at the pre-therapy phase was also not significant [$p = .85$].

7.3.2 Target-unrelated naming errors

For 'Other errors: Real word; no link to target,' an increasing trend nearing significance was detected for the 'Unseen Control' condition in P4 [$p = .054$]. The rest of the findings, including post hoc tests were not significant and/or not interpretable (Table 7.4, Appendix 7). A baseline check on the rates of making 'Other errors: Real word; no link to target' at the pre-therapy phase was also not significant [$p = .69$].

Table 7.3. Summary of Poisson main and post hoc analyses of each target-related naming error type made by each participant across various time points (pts) for initial naming responses.

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target-Related Naming Error Type / Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P1																																					
Max no. of errors at that time point	22	20	18	17	11	12		23	20	20	19	9	14		24	20	19	12	11	15		22	22	20	12	13	9		24	20	18	18	15	17			
Semantic (Sem)	1	2	2	8	2	2	n.s.	2	3	4	4	2	5	n.s.	1	2	3	2	4	2	n.s.	4	3	2	1	1	0	n.s.	2	5	2	5	3	8†	†		
Circumlocution	1	3	3	1	0	0	n.s.	2	5	3	6	1	2	n.s.	1	2	5	2	1	2	n.s.	2	2	3	1	1	0	n.s.	6	0	2	4	4	2	n.s.		
Phonological: < 50% similar	1	0	0	1	0	0	n.s.	0	0	0	1	0	0	n.s.	0	1	0	0	0	1	n.s.	0	1	0	0	0	0	n.s.	0	0	0	0	0	0	—		
Phonological: ≥ 50% similar	0	0	1	2	0	0	n.s.	0	1	0	1	0	0	n.s.	1	0	1	1	0	1	n.s.	0	1	1	0	1	0	n.s.	0	0	1	1	0	0	n.s.		
Sem/Phono.	0	0	2	1	0	2	n.s.	3	1	1	1	2	3	†	0	1	0	3	0	0	n.s.	0	1	4	1	2	1	n.s.	0	0	0	1	0	0	n.s.		
Orthographic	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	1	0	0	0	0	n.s.	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Rejection: Reject Incorrect Resp(s)	1	1	2	0	0	0	n.s.	0	0	1	2	0	0	n.s.	0	0	2	1	1	0	n.s.	0	1	3	1	0	2	n.s.	0	0	0	2	0	1	†		
P2																																					
Max no. of errors at that time point	20	19	20	20	13	11		20	19	20	20	15	16		19	20	20	6	13	14		21	21	20	16	15	15		20	19	21	18	17	19			
Semantic (Sem)	3	3	1	1	2	0	n.s.	3	2	1	2	2	5	n.s.	3	2	1	0	2	3	n.s.	1	3	1	4	3	3	†	1	2	5	2	3	5	n.s.		
Circumlocution	0	0	4	2	2	2	n.s.	0	1	2	0	1	1	n.s.	0	3	5	1	0	1	n.s.	1	4	6	2	2	1	n.s.	0	2	5	3	3	4	n.s.		
Phonological: < 50% similar	0	0	0	0	0	0	—	1	0	1	1	1	1	n.s.	0	1	0	0	1	1	n.s.	2	2	0	0	0	1	n.s.	0	0	1	0	3	1	n.s.		
Phonological: ≥ 50% similar	1	1	1	1	0	1	n.s.	0	2	0	2	1	1	n.s.	0	0	3	0	0	1	n.s.	3	0	0	0	0	0	n.s.	0	0	0	0	0	0	—		
Sem/Phono.	1	0	2	1	1	1	n.s.	0	0	1	0	2	2	*	1	3	2	2	1	1	n.s.	1	2	2	1	1	3	n.s.	0	0	0	1	1	0	n.s.		
Orthographic	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Rejection: Reject Incorrect Resp(s)	1	1	0	4	2	1	n.s.	0	2	1	2	0	1	n.s.	0	0	0	1	1	0	†	1	0	0	3	2	3	**	0	2	1	5	0	0	n.s.		

(continued)

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target-Related Error Type/ Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P3																																					
Max no. of errors at that time point	20	19	20	17	14	17		21	19	20	18	13	15		21	21	21	23	10	15		21	20	20	18	17	14		23	20	20	18	18	17			
Semantic (Sem)	1	1	2	1	1	3	<i>n.s.</i>	1	2	1	2	2	6	*	2	3	3	2	0	1	<i>n.s.</i>	1	0	2	2	1	3	<i>n.s.</i>	1	3	4	3	2	6	<i>n.s.</i>		
Circumlocution	0	1	1	1	0	1	<i>n.s.</i>	0	0	2	0	0	0	<i>n.s.</i>	1	0	0	0	0	0	<i>n.s.</i>	1	0	0	0	0	0	<i>n.s.</i>	1	1	0	0	1	2	<i>n.s.</i>		
Phonological: < 50% similar	0	0	0	0	3	0	†	1	0	2	1	0	3	<i>n.s.</i>	1	3	0	1	2	1	<i>n.s.</i>	0	1	4	1	1	1	<i>n.s.</i>	1	2	0	0	1	3	<i>n.s.</i>		
Phonological: ≥ 50% similar	0	0	0	0	0	0	—	1	0	1	2	2	0	<i>n.s.</i>	1	0	2	1	1	0	<i>n.s.</i>	1	2	2	2	5	0	<i>n.s.</i>	3	1	0	1	1	1	<i>n.s.</i>		
Sem/Phono.	1	1	1	0	0	2	<i>n.s.</i>	3	4	3	4	3	2	<i>n.s.</i>	1	0	1	3	0	1	<i>n.s.</i>	2	1	1	1	0	1	<i>n.s.</i>	1	3	2	2	3	1	<i>n.s.</i>		
Orthographic	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Rejection: Reject Incorrect Resp(s)	2	1	2	1	0	0	<i>n.s.</i>	4	3	2	0	1	0	*	2	3	5	2	1	2	<i>n.s.</i>	2	3	0	0	0	2	<i>n.s.</i>	3	2	5	0 [†]	1	0 [†]	*		
P4																																					
Max no. of errors at that time point	27	28	28	23	24	21		28	27	27	27	26	24		27	28	27	26	26	23		27	28	27	27	25	27		27	27	27	25	24	26			
Semantic (Sem)	0	2	2	1	1	1	<i>n.s.</i>	0	1	1	1	0	1	<i>n.s.</i>	1	3	2	3	0	0	*	1	2	0	1	0	0	<i>n.s.</i>	0	1	0	1	1	2	<i>n.s.</i>		
Circumlocution	0	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Phonological: < 50% similar	0	1	1	8 [†]	2	3	**	1	2	1	4	2	3	<i>n.s.</i>	0	4	4	2	5	4	<i>n.s.</i>	4	0	2	2	4	2	<i>n.s.</i>	2	1	6	2	6	4	<i>n.s.</i>		
Phonological: ≥ 50% similar	0	0	1	1	2	6	**	3	0	0	3	2	3	<i>n.s.</i>	1	0	2	1	2	3	<i>n.s.</i>	1	0	1	0	1	3	<i>n.s.</i>	2	3	3	2	4	1	<i>n.s.</i>		
Sem/Phono.	1	0	2	1	2	0	<i>n.s.</i>	0	2	1	0	1	3	<i>n.s.</i>	0	2	0	0	3	4 [†]	**	2	0	0	3	1	1	<i>n.s.</i>	2	0	4	0	0	1	<i>n.s.</i>		
Orthographic	0	0	0	0	0	0	—	0	0	0	1	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	0	0	0	0	0	—	1	0	0	0	0	0	<i>n.s.</i>		
Rejection: Reject Incorrect Resp(s)	0	0	0	3	1	0	*	1	1	1	0	2	0	<i>n.s.</i>	1	0	3	1	0	1	<i>n.s.</i>	0	2	2	1	1	0	<i>n.s.</i>	1	1	1	0	0	1	<i>n.s.</i>		

(continued)

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target-Related Error Type/ Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P5																																					
Max no. of errors at that time point	27	26	26	10	14	18		26	27	26	20	18	17		27	26	26	25	8	15		27	27	27	26	20	22		25	27	25	19	22	24			
Semantic (Sem)	5	3	4	2	1	3	<i>n.s.</i>	2	3	4	3	4	1	<i>n.s.</i>	2	1	3	3	1	2	<i>n.s.</i>	4	3	1	4	2	5	<i>n.s.</i>	2	2	1	4	3	3	<i>n.s.</i>		
Circumlocution	0	1	2	0	0	0	<i>n.s.</i>	0	1	0	1	0	0	<i>n.s.</i>	0	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Phonological: < 50% similar	1	0	0	1	0	0	<i>n.s.</i>	1	0	1	0	0	0	<i>n.s.</i>	0	0	0	0	0	1	*	2	0	0	0	0	0	<i>n.s.</i>	1	2	0	0	0	0	<i>n.s.</i>		
Phonological: ≥ 50% similar	0	0	0	1	1	0	†	0	0	1	1	1	2	<i>n.s.</i>	0	1	1	2	0	0	<i>n.s.</i>	1	0	0	0	2	0	<i>n.s.</i>	1	1	1	0	1	0	<i>n.s.</i>		
Sem/Phono.	3	0	2	2	2	2	<i>n.s.</i>	2	0	2	1	1	2	<i>n.s.</i>	2	2	1	3	1	1	<i>n.s.</i>	2	1	1	1	1	2	<i>n.s.</i>	0	1	1	0	0	2	<i>n.s.</i>		
Orthographic	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	1	0	0	<i>n.s.</i>	0	0	0	0	0	0	—		
Rejection: Reject Incorrect Resp(s)	1	0	0	1	1	2	*	0	0	0	0	1	0	<i>n.s.</i>	2	1	2	0	3	0	<i>n.s.</i>	0	2	0	0	2	2	*	0	0	1	1	1	1	<i>n.s.</i>		

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):

† .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001

- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):

† .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001

—: Not applicable for computation as values are zero across the time points

Note: Significant results are highlighted in yellow. *Significant and interpretable results* are highlighted in yellow and lettered in red (decreasing trend) or green (increasing trend).

Table 7.4. Summary of Poisson main and post hoc analyses of each target-unrelated and null-type naming error made by each participant across various time points (pts) for initial naming responses.

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target- Unrelated Naming Error Type / Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P1																																					
Max no. of errors at that time point	22	20	18	17	11	12		23	20	20	19	9	14		24	20	19	12	11	15		22	22	20	12	13	9		24	20	18	18	15	17			
Other errors: Real word; no link to target	0	0	0	0	0	0	—	1	1	0	0	0	0	<i>n.s.</i>	1	0	0	0	0	0	<i>n.s.</i>	0	0	1	0	1	0	<i>n.s.</i>	0	0	0	1	0	0	<i>n.s.</i>		
Other errors: Non-word; no link to target	0	0	0	0	0	1	†	0	1	2	0	0	0	<i>n.s.</i>	1	0	0	0	1	0	<i>n.s.</i>	0	0	1	0	0	0	<i>n.s.</i>	1	1	0	0	0	0	<i>n.s.</i>		
Other errors: Multiple words; no link to target	0	2	0	0	0	0	<i>n.s.</i>	0	0	1	0	0	0	<i>n.s.</i>	1	0	0	0	0	0	<i>n.s.</i>	1	1	0	1	0	1	<i>n.s.</i>	0	0	1	0	1	0	<i>n.s.</i>		
Miscellaneous	1	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	1	1	**	1	0	0	0	0	0	<i>n.s.</i>	1	0	1	1	1	1	<i>n.s.</i>	1	0	0	1	0	0	<i>n.s.</i>		
No response	17	11	8	4	9	7	<i>n.s.</i>	15	8	8	4	3	3	<i>n.s.</i>	17	13	8	3	4	9	<i>n.s.</i>	14	12	4	6	6	4	<i>n.s.</i>	14	14	12	3 [†]	7	6***	*		
P2																																					
Max no. of errors at that time point	20	19	20	20	13	11		20	19	20	20	15	16		19	20	20	6	13	14		21	21	20	16	15	15		20	19	21	18	17	19			
Other errors: Real word; no link to target	1	0	0	0	0	0	<i>n.s.</i>	1	0	0	1	1	0	<i>n.s.</i>	0	0	1	0	0	0	<i>n.s.</i>	0	1	0	0	0	0	<i>n.s.</i>	0	0	1	0	1	3	<i>n.s.</i>		
Other errors: Non-word; no link to target	0	0	1	0	1	1	<i>n.s.</i>	0	1	0	0	3	0	*	1	2	0	0	2	1	<i>n.s.</i>	1	3	0	2	0	1	<i>n.s.</i>	2	0	0	0	1	0	<i>n.s.</i>		
Other errors: Multiple words; no link to target	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	1	1	0	0	0	<i>n.s.</i>	0	0	1	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—		
Miscellaneous	0	0	1	0	0	0	<i>n.s.</i>	0	1	1	1	0	0	<i>n.s.</i>	1	0	0	0	0	1	<i>n.s.</i>	0	0	1	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—		
No response	13	14	10	11	5	5	<i>n.s.</i>	15	10	13	11	4	5	*	13	8	7	2	6	5	<i>n.s.</i>	11	6	9	4	7	3	<i>n.s.</i>	17	13	8	7	5	6	*		

(continued)

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target- Unrelated Error Type/ Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P3																																					
Max no. of errors at that time point	20	19	20	17	14	17		21	19	20	18	13	15		21	21	21	23	10	15		21	20	20	18	17	14		23	20	20	18	18	17			
Other errors: Real word; no link to target	0	1	3	2	2	1	<i>n.s.</i>	0	0	0	1	1	0	<i>n.s.</i>	0	2	0	2	1	4	*	0	0	1	0	1	1	<i>n.s.</i>	1	0	0	2	0	1	<i>n.s.</i>		
Other errors: Non-word; no link to target	2	1	3	7	2	6	*	6	1	1	3	2	2	<i>n.s.</i>	3	2	2	4	2	3	<i>n.s.</i>	5	5	2	4	3	4	<i>n.s.</i>	2	1	1	2	5	2	<i>n.s.</i>		
Other errors: Multiple words; no link to target	0	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	1	0	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Miscellaneous	0	0	0	1	0	1	<i>n.s.</i>	0	0	0	1	1	2	*	0	0	1	0	0	0	<i>n.s.</i>	0	0	0	0	2	0	*	0	0	0	0	0	0	—		
No response	14	12	8	4	6	3	*	5	9	8	4	1 [†]	0 ^{**}	**	10	7	7	8	3	3	<i>n.s.</i>	9	8	8	8	4	2	<i>n.s.</i>	10	7	8	8	4	1	<i>n.s.</i>		
P4																																					
Max no. of errors at that time point	27	28	28	23	24	21		28	27	27	27	26	24		27	28	27	26	26	23		27	28	27	27	25	27		27	27	27	25	24	26			
Other errors: Real word; no link to target	3	3	7	4	4	6	<i>n.s.</i>	0	4	3	4	3	2	<i>n.s.</i>	6	3	5	5	6	4	<i>n.s.</i>	0	2	4	5	3	6	<i>n.s.</i>	2	4	3	5	5	8	†		
Other errors: Non-word; no link to target	4	3	3	2	4	1	<i>n.s.</i>	6	1	5	2	4	4	<i>n.s.</i>	1	2	3	2	6	4	*	4	7	5	4	3	9	<i>n.s.</i>	1	1	2	4	1	4	<i>n.s.</i>		
Other errors: Multiple words; no link to target	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Miscellaneous	0	0	0	1	0	0	<i>n.s.</i>	0	0	0	0	0	1	<i>n.s.</i>	0	0	0	0	0	2	*	0	0	1	0	0	0	<i>n.s.</i>	0	0	0	0	0	0	—		
No response	19	18	12	2 ^{**}	8	4 [†]	***	17	16	15	12	12	7	<i>n.s.</i>	17	13	8	12	4 [†]	1 ^{**}	***	15	15	12	11	12	6	<i>n.s.</i>	16	16	8	11	7	5	†		

(continued)

	(1) Initial Letter Cue (Orthography)							(2) Repeated Exposure: Initial Letter Cue (Orthography)							(3) Whole Word Cue (Orthography)							(4) Repeated Exposure: Whole Word Cue (Orthography)							(5) Unseen Control								
Target- Unrelated Error Type/ Time Point	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Analysis across 6 time points	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Analysis across 6 time points		
P5																																					
Max no. of errors at that time point	27	26	26	10	14	18		26	27	26	20	18	17		27	26	26	25	8	15		27	27	27	26	20	22		25	27	25	19	22	24			
Other errors: Real word; no link to target	1	1	1	0	2	1	<i>n.s.</i>	1	1	0	3	4 [†]	2	**	1	2	2	2	0	2	<i>n.s.</i>	1	2	2	1	4	2	<i>n.s.</i>	1	2	1	0	0	3	<i>n.s.</i>		
Other errors: Non-word; no link to target	0	4	1	0	3	4	<i>n.s.</i>	2	3	6	2	2	3	<i>n.s.</i>	6	5	2	4	0	1	<i>n.s.</i>	3	3	1	2	2	4	<i>n.s.</i>	6	3	1	5	3	2	<i>n.s.</i>		
Other errors: Multiple words; no link to target	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—	0	0	0	0	0	0	—		
Miscellaneous	14	9	12	3	4	5	<i>n.s.</i>	14	14	11	8	4	5	<i>n.s.</i>	13	12	12	11	3	8	<i>n.s.</i>	12	9	17	15	6	6	<i>n.s.</i>	11	6	17	8	14	11	<i>n.s.</i>		
No response	2	8	4	0 [†]	0 [†]	1	*	4	5	1	1	1	2	<i>n.s.</i>	1	1	3	0	0	0	<i>n.s.</i>	2	7	5	2	1	1	<i>n.s.</i>	3	10	2	1	0	2	*		

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):

† .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001

- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):

† .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001

—: Not applicable for computation as values are zero across the time points

Note: Significant results are highlighted in yellow. *Significant and interpretable results* are highlighted in yellow and lettered in red (decreasing trend) or green (increasing trend).

7.3.3. Null-type naming errors

Several significant and interpretable results were identified for the ‘no response’ error type. For clarity, all significant findings are reported below and arranged based on the manipulated conditions. Unless stated, the remaining results, including post hoc results for ‘no response,’ were not significant. All baseline checks were also not significant (Table 7.4 and Appendix 7).

Initial letter cue (orthography). Significant trends in the decreasing direction for ‘no response’ were observed in P3 [$p = .018$] and P4 [$p < .001$]. Post hoc tests with Bonferroni’s corrections found that in P4, the decrease in the ‘no response’ error rate was significant between outcome at the final baseline versus one-week post-therapy [$p = .006$] and approached significance for the final baseline versus seven weeks post-therapy [$p = .044$].

Whole word cue (orthography). The rate of ‘no response’ showed a decreasing trend in P4’s data [$p < .001$]. Post hoc tests with Bonferroni’s corrections revealed that the decline approached significance when the error rate at the final baseline was pitted against the rate at one-week post-therapy [$p = .044$]. Additionally, the decrease was significant when the final baseline error rate was compared against the rate four weeks after therapy [$p = .002$].

Repeated exposure: Initial letter cue (orthography). Significant trends in the decreasing direction were observed in P2 [$p = .032$] and P3 [$p = .009$]. Post hoc testing with the final baseline showed that in P3, the decline was marginally significant when compared against the error rate of ‘no response’ at four weeks post-therapy [$p = .022$], while the comparison was significant against the rate yielded seven weeks later [$p = .004$], after correcting for Bonferroni’s respectively.

Repeated exposure: Whole word cue (orthography). None of the results in the five participants was significant for this control condition.

Unseen control. Significant trends in the decreasing direction were observed in P1 [$p = .015$] and P2 [$p = .022$]. A decreasing trend approaching significance was also observed in P4 [$p = .067$]. Post hoc comparisons in P1’s data, adjusted for Bonferroni’s, revealed that the decrease in the ‘no response’ error rate was marginally significant between outcome at the final baseline versus one-week post-therapy [$p = .029$], and was significant between the outcome at the final baseline versus seven weeks post-therapy [$p < .001$].

7.4 Discussion

In this chapter, data from the orthographic cueing therapy (Chapters 5 and 6) were recoded into the thirteen error types. These were in turn categorised into (1) target-related, (2) target-unrelated and (3) null-type error (Table 7.2). For each of the five conditions, error analyses were processed at the individual level using the non-parametric weighted trend test. The five conditions were the two orthographic cue conditions (initial letter and whole word cues), the two corresponding repeated exposure control conditions and the unseen control condition. Based on literature (Table 7.1), one hypothesis was formulated per error category.

Hypothesis (1): The rate of making *target-related errors* after therapy should follow an *increasing* trend; this trend is observed for items treated in the cue conditions only, but not for the control conditions.

There is no evidence in the data to support this hypothesis. Across participants, neither of the two orthographic cue conditions generated any significant increasing trends across time for any target-related errors. Although there was one increasing trend that approached marginal significance for semantic errors in one participant (P1), this result was identified in the 'Unseen control' condition (Table 7.3). This finding thus runs contrary to the hypothesis. In the absence of significant results in both cue conditions, the presence of an increasing trend of a target-related error present within a control condition suggests that a statistically significant evolution of a target-related error across time may not necessarily be driven by therapy. Reasons why error trends were not significant in the cue conditions are explored in section 7.4.3.

Hypothesis (2): The rate of making *target-unrelated errors* after therapy should follow a *decreasing* trend; this trend is observed for items treated in the cue conditions only, but not for the control conditions.

The data does not support hypothesis (2) either. Similar to hypothesis (1), none of the participants in either orthographic cue conditions produced decreasing trends for the target-unrelated errors they made. The solitary result approaching marginal significance was identified for 'Other errors: Real word; no link to target' in one participant (P4), but it was in an increasing direction. Importantly, this finding was again found in the 'Unseen control' condition (Table 7.4) – the lack of direct influence of any cues to avert participants from making unrelated errors could perhaps explain the increasing trend. Perhaps more usefully,

this reinforces how significant trend outcomes (increasing or decreasing) may not be due to treatment per se (also see section 7.4.3).

Hypothesis (3): The rate of making *null-type errors* (e.g., no response) after therapy should follow a *decreasing* trend; this trend is observed for items treated in the cue conditions only, but not for the control conditions.

There is partial support for hypothesis (3). Although significant decreasing trends were present in two participants' data for the two orthographic cue conditions (P3 and P4 for initial letter cue, and P4 for the whole word cue), these were qualified by significant trends observed in the control conditions, namely 'Repeated exposure: Initial letter cue' (P2, P3) and 'Unseen control' (P1, P2 and P4). Still, it appears that evidence for a decline in 'no response' seems stronger compared to other error types, although this decline may not necessarily be driven by either cue therapy.

7.4.1 Pooling the error analysis results together

In combination, the results suggest that changes in target-related and unrelated error types were not robust. The strongest evidence, as mentioned, appeared to be the decrease in 'no response(s).' This is consistent with other pieces of research reported in Table 7.1, e.g., K.J. in Hashimoto (2012), the group findings for treated items in Kendall et al. (2013) or group findings in Minkina et al. (2016). In these pieces of research, significant decreases in 'no response' errors were found, in contrast to the null effects identified for other error types within the same participant/dataset.

Prevalence of decreases in 'no response' from pre- to post-therapy in this dataset suggest that part of the success in orthographic cueing perhaps, lies in motivating the participants' lexical/sublexical systems into action, prompting a spoken response (word or nonword), rather than maintaining silence. It is unclear though whether these attempts in naming can be solely attributed to the written cues per se, or simply a corollary of having more encounters (opportunities) to do so, as it were with say, the repeated exposure conditions (explored in section 7.4.2).

Still, in the context of extant research (Table 7.1), the results in this chapter represent an advancement: Firstly, a more holistic statistical test was employed, in the form of non-parametric weighted trend test for Poisson distributed rates. The Poisson distribution caters for the small error sampling and by virtue of being a trend test, the analysis is also able to accommodate more data points across time. This is an advancement compared to other stroke-based aphasia error analyses that relied on comparing just two data points for main

analyses (e.g., one pre- vs one post-treatment data in Bose, 2013; paired *t*-tests in Kendall et al., 2013; Wilcoxon signed-ranks test in Minkina et al., 2019), checking effect sizes without further statistical analyses (Hashimoto, 2012) or eyeballing (Gordon, 2009). Testing across more data points mean results from the trend test is more representative in terms of capturing shifts across time, which is exactly what is measuring the trajectory of error changes is meant to accomplish. Secondly, unlike other published research, the study design here is more robust since two forms of control conditions were also incorporated. Specifically, the inclusion of repeated exposure conditions affords a better understanding of how repeated priming without any formal therapy could affect changes in naming errors. The significant outcomes in ‘no responses’ in these control conditions in at least two participants (more if the ‘unseen control’ condition is included) within this study also raises the important question of whether the error changes detected in previous studies were simply a function of repeated exposure, or were they indeed treatment effects as put forth by their authors.

7.4.2 Locus of effect for ‘repeated exposure’

Across several participants, significant changes in error rates in the control conditions were observed. Be it ‘unseen controls’ or ‘repeated exposure,’ the theme of recurrent encounters is a common thread. Probing into the possible locus may be worthwhile. This is delved deeper by examining the two ‘repeated exposure’ control conditions. Specifically, evidence of a change in error type after multiple repeated exposures is found in P2 and P3: in both, there were decreases in ‘no response’ errors in the ‘Repeated Exposure: Initial Letter Cue’ condition. The mechanism behind ‘repeated exposure’ is repetition priming. According to literature, the locus of the effect for repetition priming in picture naming can either be found in the high-level visual processing stage or in the word production stage (see review by Francis, 2014). High-level visual processing refers to the ability to recognise a different exemplar of the same item (e.g., name the same item pictured from a different perspective than the one shown during priming; Experiment 3 in Francis et al., 2008). There is no evidence of high-level visual processing in the present study, since the design of this intervention study only incorporated low-level visual processing. This is because there was only *identical repetition*, i.e., the same picture was used in the therapy phase (priming) and in post-therapy assessments. Basic visual matching of exact replicas of previously seen pictures is not considered high-level visual processing. Thus, by elimination, a more probable locus for these two participants’ data is the word production stage.

Based again on literature (e.g., Monsell, Matthews, & Miller, 1992), the priming locus in word production processes can occur anywhere from the engagement of the (a) lexical

semantic module (i.e., ‘Lexical Semantics’ in Figure 6.1, or ‘accessing the lemma (syntactic form)’ in the words of Francis, 2014 (p. 1302)) to (b) the activation of the pathway between lexical semantics and phonological output lexicon (Figure 6.1). The activation of phonological word forms in the phonological output lexicon had been ruled out, following the seminal research by Wheeldon and Monsell (1992). They ran three experiments using healthy adults and demonstrated that firstly, the locus can be in the spoken word production processes. This is because they showed that prime did not need to involve a pictorial form and could simply be a written or spoken definition of the target word, for faster object naming to be elicited as long as 12 minutes later (Experiment 1), thus ruling out (low-level) visual processes to be their locus. Secondly, their Experiments 2 and 3 showed that spoken homophone primes (same phonological word form, different meanings) did not facilitate spoken picture naming, suggesting that repeating the phonological word form in the absence of the target meaning was not sufficient to generate repetition priming effects. This means that the locus of repetition priming cannot be the activated phonological forms in the phonological output lexicon. However, the literature is ambivalent in terms of pinpointing (a) lexical semantics or (b) the pathway between lexical semantics and phonological output lexicon to be the exact locus. If the locus is in the (a) activation of the semantic system, an increase in ‘semantic’ errors will be expected across time. Conversely, if the locus is in (b) the mapping between the semantic system and the phonological output lexicon, a significant increase in ‘semantic and phonological’ mixed errors would be observed in any of the participants’ data.

Unfortunately, the data in this thesis supported neither of these two locations to be the priming locus. This is because none of the results for changes in ‘semantic’ errors and ‘semantic/phonological’ mixed errors were significant and sufficiently robust (i.e., raw values ≥ 5 for at least 50% of the time points; Table 7.3) to support either locus. The error outcomes here are therefore inconclusive in terms of adjudicating the locus for ‘repeated exposure’. To recap, in this dataset, only outcomes for ‘no response’ errors were significant and numerically reliable (Table 7.4, interpretable significant results in red). As stated in the introduction of this chapter, since there can be multiple explanations on how a ‘no response’ arises (e.g., a genuine absence of a lexical entry or that a participant self-monitors and inhibits a potential erroneous response from being produced), it is thus not possible to ascribe a locus of effect to explain the significant decreases in ‘no response(s)’ errors under ‘repeated exposure’ conditions at this stage.

7.4.3 The non-significant effects on target-related and unrelated errors following orthographic cueing

As to why the impact of orthographic cues on error changes of the 12 target-related and unrelated error types were not significant and numerically robust, there are two potential reasons. The first is that perhaps the actual influence of orthographic cues on these twelve error types may be genuinely minimal or weak. Based on the cognitive neuropsychology model (Figure 6.1), an orthographic cue could activate the semantic (path A), phonological (path B), and the orthographic modules (paths A, B and C), or a mixture of the any of these modules. This means unlike say, a semantic cue whose effect is more targeted towards the semantic system, effects of orthographic cueing might be more diffused. This also meant the impact of written cues could be less precise and correspondingly, their influence on a specific pathway and correspondingly, each specific error type, less distinct.

The second reason is logistical. As mentioned in the introduction of Chapter 7, the raw number for any given error type can be small (also see p. 654 in Gordon, 2009, on a similar point). This is because an error type may not be regularly made by a participant, or that the maximum number of opportunities to make an error is limited by the number of items treated. A slight elaboration on the latter, having multiple experimental conditions can make it challenging to have a large pool of items per condition. In this therapy study, there are as many as 120 to 140 items (exact number vary depending on the participant; numbers exclude items included for encouragement that were not analysed) distributed equally across the five conditions. These numbers also take into account what is humanely feasible for an adult with aphasia to manage. The number of treated items per cue condition and thus maximum number of opportunities to make an error for a given prompt, range from 24 to 28. Such numbers, although respectable, when spread across 12 potential errors, can mean that the sampling of each error type is small. This in turn reduces the possibility of obtaining a significant outcome per error type. Future studies can explore focusing on just one orthographic cue. Having one cue condition means it is possible to have a larger number of treated items to investigate that single cue. This could potentially drive up the error numbers generated per error type, augmenting the power of related error analyses.

Overall, results from the error analyses suggest that evidence for either initial letter or whole word orthographic cue affecting a particular error type is not strong. More importantly, the analyses clarified that although it appeared written cues could behoove participants into making fewer 'no response(s)', this may however be a non-therapy related effect since the finding is also found in the control conditions. Nonetheless, it is slightly surprising that orthographic errors were not as commonly made, despite the visual word forms of the two manipulated cues. Theoretically, having written cues could inspire participants to say, finger write out their responses, thus increasing the chances of making orthographic errors. That turned out not to be so – there was only one instance of an orthographic error (P4) observed

across the treated and untreated items and it emerged in a baseline session, so that could not be attributed to treatment either (Table 7.3). Afterall, the expected output in spoken picture-naming is to say out a whole word; if the desired outcome is for participants to be able to write out their responses (not the objective of this study), explicit instructions to do so are perhaps, necessary.

Chapter 8

General Discussion

8.1 Summary of main findings

Although estimates vary, a recently published audit on UK stroke data (2013-2015) revealed that 41% of the patients admitted for stroke were diagnosed with aphasia (Mitchell et al., 2021; also see Gronberg, Heriksson, Stenman and Lingren, 2022 on the high 30% incidence rate of aphasia in individuals with ischemic stroke based on the Swedish data collected between March 2017 and February 2018). Cumulatively, over 350,000 people within the UK have aphasia (Stroke Association, n.d.), of which spoken word-finding difficulties is a common deficit (Helm-Estabrooks & Albert, 2004). The central question posed by this thesis is therefore timely and urgent: What makes for an effective spoken word-finding therapy? Across the chapters, findings in support of orthographic cues arose from the meta-analysis of secondary data in Chapter 2 and confirmed by primary data collected from individuals with chronic anomia post-stroke.

- First, a synthesis of recent anomia therapy evidence (2008-2018) was initiated (Chapter 2). Key therapy components that underlie successful naming treatments were identified. Across short (≤ 3 weeks) and long-term (> 3 weeks) outcomes, the use of orthography within treatment (e.g., providing orthographic cues) consistently emerged as one of the top-ranked components (Figures 2.2 - 2.4).
- The meta-analytic finding on the importance of orthographic forms was put to the test in a facilitation experiment (Chapters 3 and 4). Two variations of orthographic cues (Written: initial letter and whole word) were directly pitted against their multimodal counterparts (Spoken and written: initial letter and whole word cues). Results corroborated the meta-analytic finding – At the group level, only whole word orthographic cues improved participants' final naming accuracy and naming speed (RT) over and beyond effects of repeated exposure, when assessed 15 minutes and one-week later (Table 3.6). At the individual-level, more than 50% of the sample benefitted from an "orthography-only" cue, with fewer participants benefitting significantly from the multimodal prompts (Table 4.3). Separately, correlational analyses suggested that baseline performance on a word reading test made up of both regular and exception words (CAT 20) was significantly associated with the ability to benefit from cues containing orthography.

- Finally, to adjudicate between the effectiveness of whole word and initial letter orthographic cues, the two units of written cues, in the absence of phonological components, were compared in an experimentally-controlled 13-week therapy study (Chapters 5 and 6). Overall, there was evidence for both initial letter and whole word written cues, although performances following the whole word written cues appeared more robust. At the group level, initial letter cues improved the participants' RT. The group effects of whole word cues extended beyond RT, as there was improvement on the participants' final naming accuracy as well, over and beyond effects of repeated exposure, when assessed one- and four-weeks later (Table 5.6). At the individual-level, four out of five participants significantly benefitted from the whole word orthographic cue, although two participants also profited from the initial letter prompt, with effects from initial letter lasting up to one week and whole word cues lasting as long as four weeks post-therapy (e.g., P4 and P5; see Table 6.3). Individual analyses also suggested that as long as the participants still possessed residual capabilities to process written language (i.e., they were not severely impaired on all three orthographic processing pathways shown in Figure 6.1), they stood to benefit from whole word orthographic cues. No significant changes were found in the performance of the nonword repetition control task across this time period at either the group or individual levels (Sections 5.3.3 and 6.3.2.5), suggesting any improvements observed could not be due to general improvement per se.
- Error analyses (Chapter 7) further revealed that across the 13-week therapy study, the overall improved naming accuracy could perhaps be associated to changes in the 'no response' error rates as this finding was present in all five participants, although it may not necessarily be due to cueing per se (e.g., P3 also showed similar decreasing trends making 'no response(s)' under the 'initial letter cue' condition and its corresponding 'repeated exposure' condition; also see Sections 7.4.1 and 7.4.2).

In this chapter, the main contributions from the research are discussed. This followed by more focused examination of research topics related to: (1) the mechanisms of orthographic cueing, specifically the language pathways, (2) issues of generalization, and (3) service delivery through telehealth. Throughout, suggestions on future directions are raised. Limitations and clinical implications are then discussed.

8.2 Main contribution (1): New evidence for orthographic cueing, in particular for whole word orthographic cues – Main findings and putative mechanisms

Evidence on orthographic cues in the context of therapy for spoken word-finding is important because: (1) There is a critical absence of aphasia research exclusively devoted to the role of orthography. Systematic searches for publications between January 2008 to March 2022 failed to uncover any research specific to orthography-only approaches. (2) More problematic is that there appears to be reduced appreciation of written cues, with clinicians not actively employing them (Brogan et al., 2019). Reasons for this are not clear, though the lack of direct research evidence for orthography-based approaches could have contributed to the underuse.

Evidence from this thesis can be classified into two lines; firstly, there are fresh discoveries based on published data. In Chapter 2, it was shown that orthography was a prominent predictor of participants' naming successes in multi-componential therapies previously researched (e.g., semantic features analysis in Rider et al., 2008; multimodal cueing in Snell et al., 2010). The second line of evidence will be the original data presented in Chapters 3 to 6. These included new case-level evidence in support of initial letter and whole word written cues, with effects for whole word cues lasting up to a week after a one-time application (e.g., P2 in Chapter 4) to a month after 10 sessions of orthographic cue therapy (e.g., P4 in Chapter 6). Above all, it is striking that although collected independently at different time points, the amalgamated data from literature ($n = 222$) and two sets of original data converged on the same conclusion that for participants with chronic aphasia, written cues can significantly enhance their spoken naming performances.

There are three possible reasons why orthographic cues deliver strong outcomes: (1) Visually, the print form is permanent and does not decay across time unlike the spoken format; (2) It fosters silent reading and thus phonological recoding; and (3) It triggers the motor memory involved in writing (see early evidence on this by Nakamura and colleagues, 2012), which in turn provides an additional pathway triggering successful spoken word retrieval. While these three accounts are speculative, they deserve to be investigated in future research.

Crucially, orthographic cues alone appeared sufficiently robust. Unlike orthography-only cues, outcomes from multimodal cues (orthography and phonology combined) were not as reliable, and were not quite distinguishable from effects of repeated exposure (Chapters 3 and 4). Poorer outcomes from multimodal cues could be because: (1) Receiving information from two separate modalities, the visual and auditory modalities in this case, resulted in a competition for scarce attentional resources. This competition was likely accentuated in people with post-stroke aphasia as many have comorbid attentional deficits (Murray, 2012;

also see McNeil, Odell and Tseng, 1991, for a general theory of aphasia related to the resource allocation of attention). Providing an isolated orthographic cue, however, may allow participants to dedicate their already scarce attentional resource towards the visual modality exclusively. This could result in a deeper engagement with the written cue. (2) The transient nature of the spoken cue might have exacerbated the difficulties, rendering the multimodal cue more of a distraction than a useful prompt. Participants might be compelled to allocate attention towards auditory information, and yet be unsuccessful in storing and processing the cue to aid spoken word-retrieval. Worse, this may come at the expense of not dedicating sufficient attention towards the written prompts, resulting in a “lose-lose” situation. The argument that there are deleterious effects associated with processing spoken information in a multimodal cue received preliminary empirical support in the form of negative correlations between the ability to repeat nonwords and the profiting from multimodal (orthography and phonology combined) cues. This negative finding was replicated across several outcome measures (Table 3.7). Less information, by way of involving a single modality, can deliver superior outcomes.

Turning to the family of orthographic cues, effects from whole word units appeared relatively stronger versus those of initial letters (Chapters 5 and 6), but the reasons are not entirely clear. Two plausible explanations are offered here: (1) Additional cognitive skills might be required to benefit from initial letters, making them a more difficult resource to profit from. To recap from Chapter 1, following the simplified two-stage processes inherent in all spoken word production models, there must first be accurate lexical selection followed by successful phonological access (e.g., Ellis, 1985, Dell et al., 2015, or the cognitive neuropsychological model in Figure 1.2; see Section 1.3). In lexical selection, the correct word representation within the mental dictionary (phonological output lexicon) that corresponds to the activated semantic representation is chosen. Next during phonological access, phonetic forms attached to that word representation are activated for spoken production. For successful initial letter cueing, there must be an additional capability to initiate a selection within the output lexicon narrowed down by the written initial letter prompt, to activate the correct word representation. In comparison, initiating this additional selection aided by a written letter is not required when using whole word cues. Interestingly, individual data analyses in Chapter 6 pointed to the likelihood that certain visual-based cognitive skills appeared to be required to obtain better outcomes following initial letter cues, specifically the visual-short term memory measured by ‘Corsi Block Tapping Test’ (Figure 6.5). Stronger performance on ‘Corsi Block Tapping Task’ was associated with differentiating participants who benefitted from initial letter written cues (P4 and P5) against those who did not. In contrast, participants need not demonstrate strong performances on ‘Corsi Block Tapping

Task' in order to benefit from whole word cues (e.g., P2). Exact details on how visual-short term memory may be linked to successful letter cueing are not clear at present; perhaps, as mentioned, individual differences in the level of impairment on a specialised encoding mechanism that decode letter strings within visual short-term memory could offer a plausible account (Ktori et al., 2012; see discussion on p. 196 in Chapter 6).

(2) Another explanation for the relative advantage produced by whole word orthographic cues is that whole word cues are intrinsically more informative than initial letters. Giving too little information may be inadequate to trigger the repair process and re-establish the damaged language pathways and modules. Specifying in full by way of whole words may be necessary for individuals to kickstart their language recovery process. Future research could also examine whether providing written cues up to the target words' orthographic uniqueness points would suffice, or perhaps clients genuinely require the entire word specified to optimise outcomes (e.g., 'actr' vs. 'actress').

8.3 Main contribution (2): A novel taxonomy to describe therapy components – RITA

Naming therapies are highly heterogenous. Grasping the possible permutations of 'active ingredients' within a therapy can be challenging. The meta-analysis in this thesis represents a systematic approach for decomposing these complex therapies into their distinct subparts and organise them. In the process, Chapter 2 consolidates an extensive array of therapy components into a RITA framework ('RITA': (1) Regimen; (2) Item(s); (3) Technique(s); and (4) Application of technique(s) with their Adjuncts; Appendix 2.3), which is a concrete deliverable generated by this thesis. RITA is designed as a functional and user-friendly tool, potentially helpful for students, clinicians and researchers, as they navigate the myriad of therapy options. This framework can be easily adapted for adult language therapies beyond spoken naming, as well as other range indicators, like speech or voice, or even for the paediatric population.

8.4 Main contribution (3): Applying new analytical techniques to aphasia research, while valuing data from small-*n* research designs, for fresh perspectives

The meta-analysis (Chapter 2) demonstrated how data from small-*n* case studies and experimental designs can be pooled together to productively explore a meaningful clinical question. In aphasia research, findings based on larger-scale trials are valued and evidence from small-*n* intervention studies are often dismissed (e.g., single case experimental data were excluded by Brady and colleagues' (2016) meta-analysis). This inevitably means a substantial amount of valuable clinical information is lost. A full defence of small-*n* designs in clinical research is beyond the scope of this paper. However, there are good reasons to place premium on both findings from small-*n* design studies and larger-scale clinical research, not in the least that firstly, therapists still find case studies to be useful references to gauge the existing patients on their caseloads and secondly, group averages mask idiosyncratic individual outcomes (see Section 4.1.1, or see Best, Sze, Edmundson, & Nickels, 2019, for a detailed discussion). The latter is illustrated by the differences in findings between Chapters 5 and 6. The group-level results in Chapter 5 indicated that participants only improved on their final naming accuracy and naming speed following 'Whole Word: Orthography' therapy cue, after comparing against effects from repeated exposure. It was only through individual analyses in Chapter 6 that three individuals (P2, P3 and P5) were identified as demonstrating significant gains on initial naming accuracy as well. Thus, it is important to consider individual-level outcomes; excluding results from small-*n* design research before data synthesis might unintentionally distort meta-analytic findings.

Importantly, with recent statistical advances, data accumulated from small-*n* designs can be analysed using new machine learning methods, like random forest. While this is not the first investigative attempt in anomia research to exploit the benefits offered by random forest, Chapter 2 extends the initial contribution by de Aguiar et al. (2016). These newer approaches to statistical analyses have the advantage of sidestepping problems conventional approaches wrestle with (e.g., the failure of a summary statistic to adequately address the inter-study differences, as raised by Wisenburn and Mahoney, 2009; also see Hall and Rosenthal, 1995), and are therefore capable of producing fresh insights from old data.

8.5 Main contribution (4): Highlighting the need for naming performance to be measured comprehensively

This thesis adopted a comprehensive approach towards indexing naming performance. Picture-naming appears deceptively simple; its underlying cognitive process is

in truth complex, involving different language processing stages and cognitive skills. The complexity means naming performance is correspondingly not monolithic (see pp. 1-3 in Walker, Basilakos, Fridriksson and Hickok, 2022, for a discussion). It is thus not adequately represented by a single 'right' or 'wrong' binary outcome. To this author's knowledge, this thesis represents the first time that picture-naming performance is broken down into three dimensions, namely: Initial naming accuracy, final naming accuracy and RT. This breakdown is useful because each variable captures something unique. The three variables also take on board how in everyday interaction, when an object is named, the labelling is completed quickly and accurately with minimal/nil need for self-corrections.

Past studies in aphasiology tended to focus on just one accuracy measure, i.e., either the first (e.g., Best et al., 2002; Walker et al., 2021; Hameau, Biedermann, Fieder, & Nickels, 2020) or the final naming attempt (e.g., Kendall, Pompon, Brookshire, Minkina, & Bislick, 2013; Harnish et al., 2014). However, there are strong motivations to examine both. Accuracy of the initial naming response informs us about the participants' precision from the outset. In contrast, final naming accuracy is more informative of the potential precision an individual is capable of achieving when granted more processing time. A key reason to include final naming responses is practical – individuals with unimpaired language processing monitor and self-correct their spoken utterances in daily interaction. A further reason is therapy-related —When clinicians measure naming outcomes, a primary purpose is to track a client's treatment progress (e.g., 197 out of 205 clinicians in 25 countries reported collecting aphasia outcomes to “measure progress,” see p. 1582 in Tyler et al., 2023); and if the treatment process provides the space for clients to monitor and adjust their spoken responses, the output measure should encompass these self-corrections accordingly. Thus, from a clinical standpoint, accuracies of final responses are arguably more representative of the clinical impact of the intervention.

The last variable RT measures the speed taken to produce correct responses. This is rarely investigated (Fridriksson, Richardson, Baker, and Rorden, 2011, is an exception that analysed only RT). Functionally, the promptness of lexical retrievals is an essential feature of conversation. Reducing the RT of one's response is thus a practical goal (Conroy, Drosopoulou, Humphreys, Halai, & Lambon Ralph, 2018) because dialogues will break down if one of the conversation partners takes too a long time to retrieve words. Moreover, because RT covers a naming dimension somewhat different from the correctness of spoken responses, it may capture changes in therapy progress not quantified by accuracy alone. A good example to of this is provided by P2 in the facilitation study (Chapter 4; Table 4.3). Had the research only included accuracy outcomes, a false conclusion would be drawn claiming that P2 did not benefit from whole word written cues at all, when in fact she did. P2

improved on her speed of naming following whole word cueing and this improvement was also beyond the reduced latency she achieved with multiple attempts without cues under “repeated exposure”. However, this significant piece of case-level evidence would be missed out had the analyses excluded RT. Clinicians and researchers are thus exhorted to code for the multiple dimensions of naming to avoid information loss.³⁴

Finally, if improvements are replicated beyond a single dimension, the results should then increase confidence about an intervention’s effectiveness. In the case of whole word orthographic cueing therapy, four out of the five participants improved significantly on at least two outcome measures (Table 6.3), attesting to the robustness of written prompts.

8.6 Discussion issue (1): Language pathway(s) involved in orthographic cueing – The ‘lexical-semantic account (Lorenz & Nickels, 2007) versus sub-lexical account (Best et al., 2002)’ debate

In terms of the specific language processes stimulated during orthographic cueing, there is case-level evidence from this thesis to support all three processing pathways specified in Figure 6.1. These are summarised below.

Use of the:

- Lexical-semantic pathway (path ‘A’) was evidenced by – P2, P3, and P5 (Chapter 4); P2 and P3 (Chapter 6)
- Direct lexical pathway (path ‘B’) was evidenced by – P2 (Chapter 4); P2 and P5 (Chapter 6)
- Sub-lexical pathway (path ‘C’) was evidenced by – P2 (Chapter 4); P4 (Chapter 6)

Unlike previous studies which could identify a single pathway commonly activated by all their participants to profit from a specific orthographic cue (e.g., Best et al., 2002, for the sub-lexical account, or Lorenz and Nickels, 2007, for the lexical-semantic pathway), there

³⁴ It is worth pointing out that some aphasia assessments seek to address these issues (e.g., slow responses) within their scoring systems by allocating less points to correct responses that are not prompt and require corrections. For example, the CAT subtests for naming (CAT 17 Naming Objects and CAT 18 Naming Actions), only award a full score of ‘2’ to a correct, prompt initial response, and a partial score of ‘1’ for a correct response requiring self-corrections and/or is delayed (over 5 seconds; see pp. 28-30 of the CAT manual).

was no discernible single path unifying all the participants' results in this thesis. This is perhaps not surprising, since cerebrovascular lesions do not obey the functional boundaries in a brain, causing diverse patterns of deficits. The loci of language impairments of participants were correspondingly heterogeneous, with variations in severity adding complexity. Instead of a single processing path that all participants must preserve to benefit from written cues, data from this thesis suggested that participants would recruit whichever residual path(s) still functional and capitalise on them to process the written prompts to aid word-retrieval. This is good news, because the data provides evidence that individuals will have three potential alternatives, not one processing option, to rely on to profit from orthographic cues. This perhaps explains why in chapters 4 and 6, baseline performance on a word reading task comprising both regular and irregular words (e.g., CAT 20 or PALPA 35) was associated with the rehabilitative benefits from cues containing orthographic components, since to read aloud real words would evoke either the use of one or a combination of the three orthographic input processing paths to varying extent, depending on the individual and item. However, the same association with rehabilitative benefits from orthographic cues does not extend to the reading of nonwords (e.g., PALPA 36) – Majority of the participants performed at floor on the task.

8.7 Discussion issue (2): Does orthographic cueing produce generalised improvements to untreated items?

Results from meta-analysis revealed that incorporating orthographic whole words within the therapy predicted significant improvements on untreated items, when assessed within three-weeks post-therapy (Figure 2.4). This contradicted some of the previous findings showing an absence of generalisation to untrained items following orthographic cueing (e.g., Basso et al., 2001; Miceli et al., 1996; Nickels and Best, 1996b; Lorenz and Nickels, 2007). Results from this thesis suggested that a single application of the orthographic cues would indeed not be sufficient to trigger enhanced naming of untreated items unseen during therapy (Chapters 3 and 4; Figures 4.2 - 4.4). Multiple applications of orthographic cues, however, could result in limited generalisation for a select few participants, but the effects appeared marginal: P4 and P5 in Chapter 6, for instance, showed mainly marginal improvements on final accuracy and RT when naming 'Unseen Controls' following orthographic cueing (Table 6.3 and Appendix 6.1). As for the duration, the results were variable, with marginal generalisation effects still detected for P4 seven

weeks after the end of the first therapy, whilst for P5, the carryover effects only lasted for a week (Table 6.3).

It is not clear why some individuals were able to generalise, while others did not. What was noteworthy though was that for the majority of the participants in this thesis, there was no transfer of naming benefits to the untreated (unseen control) items. This is consistent with the wider literature (Basso et al., 2001; Miceli et al., 1996) and means that outcomes stemming from orthographic cueing could inherently be limited. Despite this, it is premature to conclude so since neither of the two primary studies in this thesis manipulated the orthographic-similarity between untreated (unseen control) and the treated words. It is therefore unclear at this stage whether there was genuinely no generalization even at the word form level, or that if carryover did occur, how it is manifested. For instance, it might be possible for generalization to take place between orthographically-similar words (e.g., 'bow' to 'cow'), or perhaps it might only occur within the same word form, i.e., between homographs (e.g., 'bow' as in 'bow and tie' to 'bow' in 'bow and arrow'). Generalization at the word form level to untreated homophones (e.g., treat 'cricket' the insect improves naming 'cricket' the game) had been documented for participants with aphasia after phonological cueing therapy (Biedermann, Blanken, & Nickels, 2002; Biedermann & Nickels, 2008). Future studies on spoken naming could shed light by examining whether similar outcomes could be achieved for orthographically-related but untreated items, following orthographic cueing.

8.8 Discussion issue (3): Evidence on the feasibility of online anomia therapy

At the start of this project, there were firm plans to conduct the cueing therapies described in Chapters 5 and 6 online. This was because treatment was daily, so a practical solution was required to reduce travel logistics/cost. As mentioned in Chapter 5, this might have the unintended consequence of recruiting from households that already had relevant IT skills and were thus already well-equipped for cyber-therapies (section 5.2.3.6). Nonetheless, qualitative feedback was positive from both participants with anomia and their caregivers. On average, participants collectively rated their online therapy experience at above '5' out of the maximum '6', whether it was for whole word or initial letter orthographic cueing (Chapter 5; a rating of '6' being 'excellent'). The caregivers who assisted with daily logistics of Skype also gave an average rating above '5' on the ease of usage ('6' was the maximum and represented 'very easy'). These are encouraging figures suggesting that as long as clients have access to the appropriate IT resources at home, they are receptive

towards online word-finding treatment. Of note, almost all the participants were adults above age 65 (average age = 69.2; age range = 55 to 80) as were the three caregivers who provided IT-assistance (average age = 69.7; age range = 66 to 77), which meant that at least in the UK, age was not necessarily a barrier towards embracing e-health. Furthermore, the significant improvements achieved by four out of five participants contribute to the slender body of telehealth therapy evidence specific to anomia (see Weidner and Lowman, 2020, for a recent systematic review on overall speech pathology telehealth research; as far as this author understands, none of the language-based anomia therapy research targeting the stroke-based aphasia population included within this 2020 review, focused exclusively on orthography-based approaches).

Online language therapies bring about several advantages, aside from savings related to travelling time and cost. Covid-19 forced unusual circumstances, but created a unique opportunity to explore cross-continental online therapy. In this instance, the researcher was based in Singapore, while his last three participants resided in UK. Notably, all three participants seen by the researcher while he was overseas, significantly improved following at least one type of online orthographic cueing. Geographical boundaries, as it turned out, were not a barrier to successful anomia therapy, as long as time zone requirements were met. This implies that clients potentially do not need to be limited by constraints of their local aphasia services, especially if their immediate therapists do not provide assessments/treatments in the language that the client prefers or is dominant in. Finally, the virtual platform assures face-to-face interaction without in-person contact. This provides a sensible solution during a pandemic like Covid-19, since viral infection through human contact is high (World Health Organisation, 2020) and when the elderly are particularly at risk.

8.9 Limitations and further future directions

The two primary studies in this thesis are small-*n* case series (*n* = 5 in each). Having their findings either replicated or challenged by participants with say, different areas of impairment is particularly important, especially in the context of people with post-stroke aphasia, whose language and cognitive deficits are notoriously heterogenous (Nickels & Howard, 1995). Having further in-depth research evaluating the efficacy of orthographic whole word cues should boost clinician's confidence to apply them (Brogan et al., 2020), and

provide future researchers with a substantial body of data points to deduce more concretely the links between specific language/cognitive profiles and cue therapy outcomes.

Another potential area of future research will be to investigate if and how orthographic cues can benefit written naming. In this thesis, only spoken naming was examined. The current results support a cross-modality account, with orthographic input significantly improving a spoken output. It will be interesting to establish whether there are further benefits gleaned from orthographic cueing, including ancillary improvement within the same modality, i.e., written naming. This is plausible, given how other therapy approaches have improved outcomes in both spoken and written modalities. Copy and Recall Treatment, for instance, was originally devised to improve written naming through copying of printed words, but was found to incidentally benefit spoken naming (see case evidence of D.R., S.L., and W.D. in Besson, Rising, and Volk, 2003). Orthographic cueing for spoken picture-naming might deliver similar bimodal benefits.

Thirdly, the effects obtained for orthographic cues across the two case-series, were specific to one task, i.e., picture-naming. It is not clear whether the single-word naming gains can transfer to other spoken language activities, like daily conversation. Although qualitative ratings suggested that participants themselves perceived improvements in their overall communication ability (verbal, non-verbal; Figure 5.4, p. 159), these await verification by future research, which could include outcomes collected using picture-description or discourses. Previous research suggested that improvements in conversations are possible following related anomia treatments. Best and colleagues (2011), for example, reported individual-specific changes like the taking of more substantive turns and reduction of minimal turns, following a multimodal anomia picture-naming therapy using an orthographic and phonological task.

Finally, anomia naming performance was analysed as comprehensively as possible using measures that are coded objectively (e.g., RT) as well as qualitatively (i.e., error types). Investigation into error types was unfortunately constrained by logistical considerations, as the number per error type could be small in part due to limited item set sizes (also see section 7.4.3). Future studies could consider increasing the number of items named by focusing on just one orthographic cue, e.g., Snell et al. (2010) was able to employ a therapy set size of 60 items by focusing just on one therapy condition. This in turn could potentially increase the numbers in each error type, facilitating more robust error analyses.

8.10 Clinical implications

A primary motivation for this thesis is to identify ways to optimise word-finding therapies. To conclude, these are some practical take-aways from this thesis.

(1) Application of whole word written cues: Where appropriate, clinicians could consider employing orthographic whole word cues during spoken word-finding therapy. Empirical evidence for this, from the meta-analysis and two primary research studies, is strong (e.g., it significantly benefitted 80% of the sample in Chapter 6). Qualitatively, four out of five participants also favoured whole word written cueing, stating that it was not frustrating unlike say, the initial letter cues (Section 5.3.1.2).

To screen for suitable clients, assessments reliant on visual word stimuli, like the visual lexical decision task and the single word reading task could be useful. The latter should comprise both regular and exception words; CAT 20 or PALPA 35, administered in Chapters 3 and 5, are two such standardised word reading tests. More research is required to establish the exact battery of orthography-based tasks relevant for screening. However, individuals with severe difficulties processing visually-presented words across tasks, including the reading aloud of words, may not be good candidates for an orthographic cue therapy (e.g., P1 in Chapter 6). Instead, for them, alternative treatment approaches like phonological cueing (Biederman, Blaken, & Nickels, 2002), and/or teaching of compensatory strategies (Russo et al., 2017) could be considered.

Importantly, results from the facilitation study challenge the mindset of administering multimodal phonological and orthographical cueing for all clients with anomia. Coelho, Sinotte, and Duffy (2008) summarised this school of thought neatly, stating that "...multimodality stimulation can improve response adequacy for many patients with aphasia, and combining auditory and visual stimulation may be the best and most practical way of doing so" (pp. 413-414). This is perhaps why cueing hierarchies in popular commercial aphasia therapy apps, like Tactus Naming Therapy (Tactus Therapy Solutions Ltd., 2022), are unless adjusted, set by default to the simultaneous presentation of phonological and orthographic cues. It is perhaps helpful to be reminded that treatments ought to be tailored to individual clients, rather than a blind adherence to a blanket doctrine.

(2) Choosing functional target words in treatment: results from the two primary research studies showed that for majority of the participants, effects from orthographic cueing were item-specific. Given the primacy of item-specific outcomes, it is important to choose personally-relevant words as therapy targets when applying orthographic cueing (see for example pp. 995-996 in Hickin et al., 2002, on a similar point). This maximises the likelihood that any successfully retrieved words can be directly employed by the clients when naming is required.

(3) Treatment schedule: On average, effects from a single application of a whole word orthographic cue could potentially last for a week (Chapter 3). After ten sessions (about half hour each; approximately five hours of therapy in total), average effects were marginal after 4 weeks (Chapter 5), although for some participants (e.g., P4 and P5 in Chapter 6), improvements remained significant 4 weeks later. The treatment regimen used by the therapy study (Chapter 5) aligns with the recent network analysis of archival data by the REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators (2022). Their results suggested that maximum gains for picture naming was associated with up to 5 hours of total therapy time (Figure 4D, p. 962), which is in keeping with the regimen specified for Chapter 5's therapy study.³⁵ Since on average, effects from whole word orthographic cueing in this thesis appeared to wane with participants' gains becoming marginal when measured a month later, perhaps clinicians could plan a booster session about three weeks post-therapy to optimise maintenance. Treatments can be administered online to ease logistics of multiple sessions.

(4) Finally, echoing the earlier suggestions for future researchers, it may be worth clinicians measuring other aspects of naming too, coding beyond the typical final naming accuracy. The need to account for these multiple dimensions is also acknowledged by picture-naming assessments that only credit a full score if the picture is named correctly without delay or correction (e.g., CAT 17 and 18 as previously described). If the concern is the additional clinical effort expended on data extraction particularly for RTs, this latency outcome can be roughly gauged using stopwatches or related phone apps in the clinic. Additional outcome information produces a more complete picture of the client's therapy progress. For the clients, it can be heartening or even inspiring to learn of any improvement they achieved, especially if it is different from the accuracy performance they are typically assessed on.

In summary, this thesis shines a spotlight on the important role orthographic cues perform. Cueing therapy has been a traditional mainstay in speech and language therapists' arsenal (Linebaugh & Lehner, 1977; Raymer & Roitsch, 2023). However, its process has evolved, with the cue composition and steps altered variously by different clinicians and researchers (Patterson, 2001). Heavier emphasis in both practice and research has always been tilted towards multimodal, semantic or phonology-only cues. Relegated too often to a

³⁵ The confidence interval reported by RELEASE Collaborators for this result is large (1.34 to 23.62). This is likely in part a consequence of combining heterogenous aphasia studies to derive a summary effect size – a limitation of conventional meta-analytic approaches as mentioned in Chapter 2 (p. 32). Their networks for picture naming constructed by the archival data appeared sparse too (e.g., Figure 2 for dosage in RELEASE Collaborators, 2022); it is not clear what implications this sparseness has on their meta-analysis findings.

second fiddle, written-only cues deserve dedicated scientific attention by aphasiologists. In this thesis, empirical evidence for whole word orthographic cues were supplied using both large-*n* secondary data as well as group and individual-level analyses of primary data from the facilitation and therapy studies. It is hoped that incorporating both large-scale and case-series designs address the contradictory concerns that: (a) On one hand, researchers are advised to study beyond individual case studies to draw convincing conclusions (see pp. 49 and 50 in Brady et al., 2016), (b) Yet, clinicians are aware that outcomes derived from groups of people of aphasia are not clinically meaningful since there is no “average” individual with aphasia (Nickels & Howard, 1995). Remarkably, whether analyses were based on large-*n* machine learning, small-*n* group averages or idiosyncratic individual-level data, in this thesis, the robust effects from orthographic whole word cues were successfully replicated across different sets of data. If the pen is mightier than the sword, then perhaps written words even in the palest ink should not be wasted, to maximise anomia rehabilitation gains.

References

- Abel, S., Weiller, C., Huber, W., & Willmes, K. (2014). *Neural underpinnings for model-oriented therapy of aphasic word production*. *Neuropsychologia*, 57, 154-165.
<https://doi.org/10.1016/j.neuropsychologia.2014.03.010>
- Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology*, 21(5), 411-447.
<https://doi.org/10.1080/02687030701192687>
- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814-823. <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). The Hatfield Image Test (HIT): A new picture test and norms for experimental and clinical use. *Journal of Clinical and Experimental Neuropsychology*, 31(6), 731-753.
<https://doi.org/10.1080/13803390802488103>
- Adrian, J. A., Gonzalez, M., Buiza, J. J., & Sage, K. (2011). Extending the use of Spanish Computer-assisted Anomia Rehabilitation Program (CARP-2) in people with aphasia. *Journal of Communication Disorders*, 44(6), 666-677.
<https://doi.org/10.1016/j.jcomdis.2011.06.002>
- Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Audacity Team (2018). Audacity (Version 2.3.0) [Computer software].
<https://www.audacityteam.org/>
- Altmann, L. J. P., Hazamy, A. A., Carvajal, P. J., Benjamin, M., Rosenbek, J. C., & Crosson, B. (2014). Delayed stimulus-specific improvements in discourse following anomia treatment using an intentional gesture. *Journal of Speech, Language, & Hearing Research*, 57(2), 439-454. [https://doi.org/10.1044/1092-4388\(2013/12-0224\)](https://doi.org/10.1044/1092-4388(2013/12-0224))

- Avila, C., Lambon Ralph, M. A., Parcet, M-A., Geffner, D., & Gonzalez-Darder, J-M. (2001). Implicit word cues facilitate impaired naming performance: Evidence from a case of anomia. *Brain and Language*, 79(2), 185-200. <https://doi.org/10.1006/brln.2001.2472>
- Balota, D. A., Yap, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459. <https://doi.org/10.3758/BF03193014>
- Barr, P., Biedermann, B., Tainturier, M-J., Kohnen, S., & Nickels, L. (2020). Too harts, won sole: Using dysgraphia treatment to address homophone representation. *Neuropsychological Rehabilitation*, 30(10), 2035-2066. <https://dx.doi.org/10.1080/09602011.2019.1629302>
- Basso, A., Marangolo, P., Piras, F., & Galluzzi, C. (2001). Acquisition of new words in normal subjects: A suggestion for the treatment of anomia. *Brain and Language*, 77(1), 45-59. <https://doi.org/10.1006/brln.2000.2422>
- Beeke, S., Beckley, F., Johnson, F., Heilemann, C., Edwards, S., Maxim, J., & Best, W. (2015). Conversation focused aphasia therapy: Investigating the adoption of strategies by people with agrammatism. *Aphasiology*, 29(3), 355-377, DOI: 10.1080/02687038.2014.881459
- Beeson, P. M., Rising, K., & Volk, J. (2003). Writing treatment for severe aphasia: Who benefits? *Journal of Speech, Language, & Hearing Research*, 46(5), 1038-1060. [https://doi.org/10.1044/1092-4388\(2003/083\)](https://doi.org/10.1044/1092-4388(2003/083))
- Beeson, P. M. & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, 16(4), 161-169. <https://dx.doi.org/10.1007/s11065-006-9013-7>
- Berry, K. J., Johnston, J. E., & Mielke, Jr., P. W. (2007). An alternative measure of effect size for Cochran's Q test for related proportions. *Perceptual and Motor Skills*, 104(3 Supplement), 1236-1242. <https://doi.org/10.2466/pms.104.4.1236-1242>
- Best, W., Grassly, J., Greenwood, A., Herbert, R., Hickin, J., & Howard, D. (2011). A controlled study of changes in conversation following aphasia therapy for anomia.

Disability & Rehabilitation, 33(3), 229-242.
<https://dx.doi.org/10.3109/09638288.2010.534230>

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex*, 49(9), 2345-2357. <https://doi.org/10.1016/j.cortex.2013.01.005>

Best, W., Greenwood, A., Grassly, J., & Hickin, J. (2008). Bridging the gap: can impairment-based therapy for anomia have an impact at the psycho-social level? *International Journal of Language & Communication Disorders*, 43(4), 390-407.
<https://doi.org/10.1080/13682820701608001>

Best, W., Herbert, R. Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, 16(1-2), 151-168. <https://doi.org/10.1080/02687040143000483>

Best, W., Maxim, J., Heilemann, C., Beckley, F., Johnson, F., Edwards, S. I., Howard, D. & Beeke, S. (2016). Conversation therapy with people with aphasia and conversation partners using video feedback: A group and case series investigation of changes in interaction. *Frontiers in Human Neuroscience*, 10, 562.
<https://doi.org/10.3389/fnhum.2016.00562>

Best, W., Nickels, L., & Williamson, L. (2005). A coding system for picture naming responses with a focus on semantic errors. [Unpublished manuscript]. Department of Language and Cognition, University College London.

Best, W., Sze, W. P., Edmundson, A., & Nickels, L. (2019). What counts as evidence? Swimming against the tide: Valuing both clinically informed experimentally controlled case series and randomized controlled trials in intervention research. *Evidence-Based Communication Assessment and Intervention*, 13(3), 107-135.
<https://doi.org/10.1080/17489539.2019.1597444>

Beukelman, D. R., & Mirenda, P. (Eds.). (2013). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (4th ed.). Chelsea, Michigan: Paul H. Brookes Publishing Co, Inc.

- Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, 34(4), 987-993. <https://doi.org/10.1161/01.STR.0000062343.64383.D0>
- Biedermann, B., Blanken, G., & Nickels, L. (2002). The representation of homophones: Evidence from remediation. *Aphasiology*, 16(10-11), 1115-1136. <https://doi.org/10.1080/02687030244000545>
- Biedermann, B., Fieder, N., & Nickels, L. (2018). Spoken word production: Processes and potential breakdown. In A. Bar-On & D. Ravid (Eds.). *Handbooks of Applied Linguistics. Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 155-177). Leck, Germany: De Gruyter Mouton. <https://doi.org/10.1515/9781614514909>
- Biedermann, B. & Nickels, L. (2008). The representation of homophones: More evidence from the remediation of anomia. *Cortex*, 44(3), 276-293. <https://doi.org/10.1016/j.cortex.2006.07.004>
- Biel, M., Nitta, L., & Jackson, C. (2018). Understanding motivation in aphasia rehabilitation. In Coppens, P., & Patterson, J. (Eds.), *Aphasia rehabilitation: Clinical challenges* (pp. 393-495). Burlington, MA: Jones & Barlett Learning.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex: John Wiley & Sons, Ltd.
- Bose, A. (2013). Phonological therapy in jargon aphasia: Effects on naming and neologisms. *International Journal of Language and Communication Disorders*, 48(5), 582-595. <https://doi.org/10.1111/1460-6984.12038>
- Boswell, M. T. (1966). Estimating and testing trend in a stochastic process of Poisson type. *The Annals of Mathematical Statistics*, 37(6), 1564-1573. <https://www.jstor.org/stable/2238773>
- Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation*, 17(6), 411-422. <https://doi.org/10.1310/tsr1706-411>

- Boyle, M. (2017). Semantic treatments for word and sentence production deficits in aphasia. *Seminars in Speech and Language*, 38(1), 52-61. <https://doi.org/10.1055/s-0036-1597256>
- Boyle, M. & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech Language Pathology*, 4(4), 94-98. <https://doi.org/10.1044/1058-0360.0404.94>
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, 6(CD000425). <https://doi.org/10.1002/14651858.CD000425.pub4>
- Breiman, L. (1999). *Random forest – Random features*. (Technical Report 567). <https://www.stat.berkeley.edu/~breiman/random-forests.pdf>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5), e10773. <https://doi.org/10.1371/journal.pone.0010773>
- Brogan, E., Godecke, E., & Ciccone, N. (2020). Behind the therapy door: what is “usual care” aphasia therapy in acute stroke management? *Aphasiology*, 34(10), 1291-1313. <https://doi.org/10.1080/02687038.2020.1759268>
- Bruce, C. & Howard, D. (1987). Computer-generated phonemic cues: An effective aid for naming in aphasia. *British Journal of Disorders of Communication*, 22(3), 191-201. <https://doi.org/10.3109/13682828709019862>
- Bruce, C. & Howard, D. (1988). Why don't Broca's aphasics cue themselves? An investigation of phonemic cueing and tip of the tongue information. *Neuropsychologia*, 26(2), 253-264. [https://doi.org/10.1016/0028-3932\(88\)90078-4](https://doi.org/10.1016/0028-3932(88)90078-4)

- Brysbaert, M., & 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*, 41(4), 977-990. <https://doi.org/10.3758/BRM.41.4.977>
- Bureau, A., Dupuis, J., Falls, K., Lunetta, K. L., Hayward, B., Keith, T. P., & Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology*, 28(2), 171-182. doi: 10.1002/gepi.20041
- Burgess, P. W. & Shallice, T. (1997). The Hayling and Brixton Tests. Pearson: London, U. K.
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.). *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, N. J.: Lawrence Erlbaum Associates, Inc.
- Caramazza, A. & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, 26(1), 95-122. [https://doi.org/10.1016/s0010-9452\(13\)80077-9](https://doi.org/10.1016/s0010-9452(13)80077-9)
- Caramazza, A. & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, 349, 788-789. <https://doi.org/10.1038/349788a0>
- Casarin, F. S., Branco, L., Pereira, N., Kochhann, R., Gindri, G., & Fonseca, R. P. (2014). Rehabilitation of lexical and semantic communicative impairments: An overview of available approaches. *Dementia & Neuropsychologia*, 8(3), 266-277. <https://doi.org/10.1590/S1980-57642014DN83000011>
- Cavanaugh, R., Quique, Y. M., Dickey, M. W., Hula, W. D., Boss, E., & Evans, W. S. (in print). Practice-Related Predictors of Semantic Feature Verification Treatment for Aphasia. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/2021_AJSLP-21-00296
- Chomeya, R. (2010). Quality of psychology test between likert scale 5 and 6 points. *Journal of Social Sciences*, 6(3), 399-403. <https://doi.org/10.3844/jssp.2010.399.403>
- Claessen, M. H. G., van der Ham, I. J., M., & van Zandvoort, M. J. E. (2015). Computerization of the standard Corsi Block-tapping Task affects its underlying cognitive

- concepts: A pilot study. *Applied Neuropsychology: Adult*, 22(3), 180-188.
<https://doi.org/10.1080/23279095.2014.892488>
- Code, C. & Tesak, J. (2008). *Milestones in the history of aphasia: Theories and protagonists*. Hove, East Sussex: Psychological Press.
- Coelho, C. A., Sinotte, M. P., & Duffy, J. R. (2008). Schuell's stimulation approach to rehabilitation. In Roberta Chapey (Ed.). *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed.) (pp. 403-449). Baltimore, Maryland: Wolters Kluwer, Lippincott Williams & Wilkins.
- Coltheart, M. (2012). Dual-route theories of reading aloud. In James S. Adelman (Ed.). *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 3-27). Hove, East Sussex: Psychological Press.
- Conroy, P., Drosopoulou, C. S., Humphreys, G. F., Halai, A. D., & Lambon Ralph, M. A. (2018). Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. *Brain*, 141(6), 1815-1827.
<https://doi.org/10.1093/brain/awy087>
- Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009a). A comparison of word versus sentence cues as therapy for verb naming in aphasia. *Aphasiology*, 23(4), 462-482.
<https://doi.org/10.1080/02687030802514920>
- Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009b). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*, 23(11), 1311-1337.
<https://doi.org/10.1080/02687030902756439>
- Conroy, P., Sage, K., & Lambon Ralph, M. (2009c). Improved vocabulary production after naming therapy in aphasia: Can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders*, 44(6), 1036-1062.
<https://doi.org/10.3109/13682820802585975>
- Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009d). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology*, 23(6), 707-730. <https://doi.org/10.1080/02687030802165574>

- Conroy, P. & Scowcroft, J. (2012). Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study. *Neuropsychological Rehabilitation*, 22(2), 295-318. <https://doi.org/10.1080/09602011.2011.641434>
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research*, 169. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Creet, E., Morris, J., Howard, D., & Nickels, L. (2019). Name it again! Investigating the effects of repeated naming attempts in aphasia. *Aphasiology*, 33(10), 1202-1226. <https://doi.org/10.1080/02687038.2019.1622352>
- Croft, S., Marshall, J., Pring, T., & Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International Journal of Language and Communication Disorders*, 46(1), 48-62. <https://doi.org/10.3109/13682822.2010.484845>
- Crofts, B., Nickels, L., Makin, K., Tylor, C., & Moss, M. (2004, August 29 - September 2). *A comparison of the effectiveness of 'semantic' and 'phonological' tasks in the facilitation of word production in aphasia* [Conference presentation]. XXVIth World Congress of the International Association of Logopedics and Phoniatrics, Brisbane, Australia.
- Dabul, B. (2000). *Apraxia Battery for Adults* (2nd ed.). Austin, Texas: Pro-Ed.
- Davis, C.H., Farias, D. & Baynes, K. (2005) Understanding the effects of cuing strategies through error analysis. *Brain and Language*, 95(1), 183-184. <https://doi.org/10.1016/j.bandl.2005.07.098>
- de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving production of treated and untreated verbs in aphasia: A meta-analysis. *Frontiers in Human Neuroscience*, 10(468). <https://doi.org/10.3389/fnhum.2016.00468>
- de Jong-Hagelstein, M., van de Sandt-Koenderman, W. M. E., Prins, N. D., Dippel, D. W. J., Koudstaal, P. J., & Visch-Brink, E. G. (2011). Efficacy of early cognitive-linguistic treatment and communicative treatment in aphasia after stroke: A randomised controlled trial (RATS-2). *Journal of Neurology, Neurosurgery & Psychiatry*, 82(4), 399-404. <https://doi.org/10.1136/jnnp.2010.210559>

- Dean, M. P. (2010). Use of orthography in spoken naming in aphasia: A case study. *Cognitive and Behavioral Neurology*, 23(4), 262-268.
<https://doi.org/10.1097/WNN.0b013e3181e078ca>
- Dechene, L., Tousignant, M., Boissy, P., Macoir, J., Heroux, S., Hamel, M., Briere, S., & Page, C. (2011). Simulated in-home teletreatment for anomia. *International Journal of Telerehabilitation*, 3(2), 3-10. <https://doi.org/10.5195/IJT.2011.6075>
- Dell, G. S., Nozari, N., & Oppenheim, G. M. (2015). Word production: Behavioral and computational considerations. In Matthew Goldrick, Victor S. Ferreira, & Michele Miozzo (Eds.). *Oxford library of psychology. The Oxford handbook of language production* (pp. 88-104). New York, N.Y.: Oxford University Press.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4), 801-838.
<https://doi.org/10.1037/0033-295x.104.4.801>
- DeLong, C., Nessler, C., Wright, S., & Wambaugh, J. (2015). Semantic feature analysis: Further examination of outcomes. *American Journal of Speech-Language Pathology*, 24(4), S864-879. doi: 10.1044/2015_AJSLP-14-0155
- Di Carlo, L. M. (1980). Language recovery in aphasia: Effect of systematic filmed programmed instruction. *Archives of Physical Medicine and Rehabilitation*, 61(1):41-44.
- Dignam, J., Copland, D., O'Brien, K., Burfein P., Khan, A., & Rodriguez, A. D. (2017). Influence of cognitive ability on therapy outcomes for anomia in adults with chronic poststroke aphasia. *Journal of Speech, Language, and Hearing Research*, 60(2), 406-421. https://doi.org/10.1044/2016_JSLHR-L-15-0384
- Døli, H., Helland, W. A., Helland, T., & Specht, K. (2021). Associations between lesion size, lesion location and aphasia in acute stroke. *Aphasiology*, 35(6), 745-763.
<https://doi.org/10.1080/02687038.2020.1727838>
- Druks, J., & Masterson, J. (2000). *An Object and Action Naming Battery*. Hove, East Sussex: Psychology Press.

- Edmonds, L. A. (2014). Tutorial for Verb Network Strengthening Treatment (VNeST): Detailed description of the treatment protocol with corresponding theoretical rationale. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, 24(3), 78-88. <https://doi.org/10.1044/nnsld24.3.78>
- Ellis, A. W. (1985). The production of spoken words: A cognitive neuropsychological perspective. In Andrew W. Ellis (Ed.). *Progress in the psychology of language: Vol. 2. Progress in the psychology of language* (pp. 107-145). London, U.K.: Lawrence Erlbaum Associates Ltd., Publishers.
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005). Treatment of anomia using errorless versus errorful learning: Are frontal executive skills and feedback important? *International Journal of Language & Communication Disorders*, 40(4): 505-523. <https://doi.org/10.1080/13682820500138572>
- Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, 16(2), 129-154. <https://doi.org/10.1080/09602010443000254>
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43(2), 182–216. <https://doi.org/10.1006/jmla.2000.2716>
- Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic Bulletin & Review*, 21(5), 1301-1308. <https://doi.org/10.3758/s13423-014-0610-9>
- Francis, W. S., Corral, N. I., Jones, M. L., & Sáenz, S. P. (2008). Decomposition of repetition priming components in picture naming. *Journal of Experimental Psychology: General*, 137(3), 566-590. <https://doi.org/10.1037/0096-3445.137.3.566>
- Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1), 20-33. https://www.ijis.net/ijis5_1/ijis5_1_freelon.pdf
- Freelon, D. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science*, 8(1), 10-16. https://www.ijis.net/ijis8_1/ijis8_1_freelon.pdf

- Fridriksson, J., Baker, J. M., & Moser, D. (2009). Cortical mapping of naming errors in aphasia. *Human Brain Mapping*, 30(8), 2487-2498. <https://doi.org/10.1002/hbm.20683>
- Fridriksson, J., Richardson, J. D., Baker, J. M., & Rorden, C. (2011). Transcranial direct current stimulation improves naming reaction time in fluent aphasia: A double-blind, sham-controlled study. *Stroke*, 42(3), 819-821. <https://doi.org/10.1161/STROKEAHA.110.600288>
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47(1), 27-52. <https://doi.org/10.2307/412187>
- Georgieff, N., Dominey, P. F., Michel, F., Marie-Cardine, M., & Dalery, J. (1998). Anomia in major depressive state. *Psychiatry Research*, 77(3), 197–208. [https://doi.org/10.1016/S0165-1781\(98\)00008-0](https://doi.org/10.1016/S0165-1781(98)00008-0)
- Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination (BDAE-3)* (3rd ed.). Philadelphia, Pennsylvania: Lippincott Williams & Wilkins.
- Gordon, J. K. (2007). A contextual approach to facilitating word retrieval in agrammatic aphasia. *Aphasiology*, 21(6-8), 643-657. <https://doi.org/10.1080/02687030701192141>
- Green, S. B., & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analysing and understanding data* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: A movement in crisis? *BMJ*, 348, g3725. <https://doi.org/10.1136/bmj.g3725>
- Grönberg, A., Henriksson, I., Stenman, M., & Lindgren, A. G. (2022). Incidence of aphasia in ischemic stroke. *Neuroepidemiology*, 56(3), 174-182. <https://doi.org/10.1159/000524206>
- Grossman, M. (2014). Anomia. In M. J. Aminoff, R. B. Daroff (Eds.). (2014). *Encyclopedia of the neurological sciences: Volume 1* (2nd ed.) (pp. 194-197). San Diego, CA: Academic Press.
- Hall, J. A. & Rosenthal, R. (1995). Interpreting and evaluating meta-analysis. *Evaluation & The Health Professions*, 18(4), 393-407. <https://doi.org/10.1177/016327879501800404>

- Harley, T. A., Oliver, T. M., Jessiman, L. J., & MacAndrew, S. B. G. (2013). Ageing makes us dyslexic. *Aphasiology*, 27(4), 490-505. <https://doi.org/10.1080/02687038.2013.775564>
- Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., Gonzalez Rothi, L. J., & Crosson, B. (2014). Dosing of a cued picture naming treatment for anomia. *American Journal of Speech-Language Pathology*, 23(2), S285-S299. https://doi.org/10.1044/2014_AJSLP-13-0081
- Hartung, J., Knapp, G., & Sinha, B. K. (2008). *Statistical meta-analysis with applications*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Hameau, S., Biedermann, B., Fieder, N., & Nickels, L. (2020). Investigation of the effects of semantic neighbours in aphasia: A facilitated naming study. *Aphasiology*, 34(7), 840-864. <https://doi.org/10.1080/02687038.2019.1652241>
- Hashimoto, N. (2012). The use of semantic- and phonological-based feature approaches to treat naming deficits in aphasia. *Clinical Linguistics & Phonetics*, 26(6), 518-553. <https://doi.org/10.3109/02699206.2012.663051>
- Hashimoto, N. (2016). The use of one or three semantic associative primes in treating anomia in aphasia. *American Journal of Speech-Language Pathology*, 25(4S), S665-S686. https://doi.org/10.1044/2016_AJSLP-15-0085
- Hatfield, F. M., & Zangwill, O. L. (1974). Ideation in aphasia: The picture-story method. *Neuropsychologia*, 12, 389-393.
- Helm-Estabrooks, N. & Albert, M. L. (2004). *Manual of aphasia and aphasia therapy* (2nd ed.). Austin, Texas: Pro-Ed.
- Hendricks, C. T., Nicholas, M. L., & Zipse, L. (2014). Effects of phonological neighbourhood on the treatment of naming in aphasia. *Aphasiology*, 28(3), 338-358. <https://doi.org/10.1080/02687038.2013.866209>
- Hermansen, S. W. (2008). Evaluating predictive models: Computing and interpreting the c statistic. <https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/143-2008.pdf>

- Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology*, 16(10/11), 981-999.
<https://doi.org/10.1080/02687030244000509>
- Hoffmann, T., Glasziou, P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods M, McCulloch, P., Wyatt, J. C., Chan A-W., & Michie S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348: g1687. <https://doi.org/10.1136/bmj.g1687>
- Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2019). *R Package 'party': A laboratory for recursive partitioning* (Version 1.3-3) [Computer software]. The R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/party/index.html>
- Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*, 29(5), 526-562.
<https://doi.org/10.1080/02687038.2014.985884>
- Howard, D. & Harding, D. (1998). Self-cueing of word retrieval by a woman with aphasia: Why a letter board works. *Aphasiology*, 12(4-5), 399-420.
<https://doi.org/10.1080/02687039808249540>
- Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). Treatment of word retrieval deficits in aphasia. A comparison of two therapy methods. *Brain*, 108(4), 817-829. <https://doi.org/10.1093/brain/108.4.817>
- Howard, D. & Patterson, K. (1992). *The Pyramids and Palm Trees Test*. London, UK: Harcourt Assessment.
- Howell, D. C. (2014). *Fundamental statistics for the behavioural sciences* (8th ed.). Belmont, CA: Wadsworth Cengage Learning.
- Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlation analysis. *Journal of Educational Psychology*, 70(6), 873-876.
<https://doi.org/10.1037/0022-0663.70.6.873>

- Icht, M. & Ben-David, B. M. (2014). Oral-diadochokinesis rates across languages: English and Hebrew norms. *Journal of Communication Disorders*, 48, 27-37.
<https://doi.org/10.1016/j.jcomdis.2014.02.002>
- Jacquemot, C., Dupoux, E., Robotham, L., & Bachoud-Lévi, A-C. (2012). Specificity in rehabilitation of word production: A meta-analysis and a case study. *Behavioural Neurology*, 25(2), 73-101. <https://doi.org/10.3233/BEN-2012-0358>
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia, Pennsylvania: Lippincott Williams & Wilkins.
- Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)*. Hove, East Sussex: Lawrence Erlbaum Associates Ltd.
- Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic Assessments of Language Processing in Aphasia (PALPA): An introduction. *Aphasiology*, 10(2), 159-180.
<https://doi.org/10.1080/02687039608248403>
- Kendall, D. L., Pompon, R. H., Brookshire, C. E., Minkina, I., & Bislick, L. (2013). An analysis of aphasic naming errors as an indicator of improved linguistic processing following phonomotor treatment. *American Journal of Speech-Language Pathology*, 22(2), S240-S249. [https://doi.org/10.1044/1058-0360\(2012/12-0078\)](https://doi.org/10.1044/1058-0360(2012/12-0078))
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81-93.
<https://doi.org/10.2307/2332226>
- Kertesz, A. (2007). *Western Aphasia Battery Revised*. San Antonio, TX: Pearson.
- Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research*, 51(6), 1550-1568. [https://doi.org/10.1044/1092-4388\(2008/07-0038\)](https://doi.org/10.1044/1092-4388(2008/07-0038))
- Kiran, S., & Johnson, L. (2008). Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech-Language Pathology*, 17(4), 389-400. [https://doi.org/10.1044/1058-0360\(2008/06-0085\)](https://doi.org/10.1044/1058-0360(2008/06-0085))

- Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language*, 24(2), 266–283. [https://doi.org/10.1016/0093-934X\(85\)90135-X](https://doi.org/10.1016/0093-934X(85)90135-X)
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Kristinsson, S., Basilakos, A., Elm, J., Spell, L. A., Bonilha, L., Rorden, C., den Ouden, D. B., Cassarly, C., Sen, S., Hillis, A., Hickok, G., & Fridriksson, J. (2021). Individualized response to semantic versus phonological aphasia therapies in stroke. *Brain Communications*, 3(3), fcab174. <https://doi.org/10.1093/braincomms/fcab174>
- Ktori, M., Grainger, J., & Dufau, S. (2012). Letter string processing and visual short-term memory. *Quarterly Journal of Experimental Psychology*, 65(3), 465-473. <https://doi.org/10.1080/17470218.2011.611889>
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1-26. <https://doi.org/10.18637/jss.v028.i05>
- Kuhn, M. (2018). *caret (Classification And Regression Training) R package* [Version 6.0-81] (Computer software). <https://github.com/topepo/caret/>
- Laine, M., & Martin, M. (2006). *Anomia: Theoretical and clinical aspects*. New York, NY: Psychology Press.
- Lam, J. M. C. & Wodchis, W. P. (2010). The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents. *Medical Care*, 48(4), 380-387. <https://doi.org/10.1097/MLR.0b013e3181ca2647>
- Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (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), 289-305. <https://doi.org/10.1080/09602010903237875>
- Lavoie, M., Bier, N., & Macoir, J. (2019). Efficacy of a self-administered treatment using a smart tablet to improve functional vocabulary in post-stroke aphasia: A case-series study.

- International Journal of Language & Communication Disorders*, 54(2), 249-264.
<https://doi.org/10.1111/1460-6984.12439>
- Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology*, 22(9), 923-947. <https://doi.org/10.1080/02687030701831474>
- Leustek, J. (2017). Contrast analysis. In Mike Allen (Ed.). *The SAGE encyclopedia of communication research methods: Vol 1* (pp. 248-252). Thousand Oaks, CA: SAGE Publishers, Inc.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, 22(1), 1-38.
<https://doi.org/10.1017/s0140525x99001776>
- Liaw, A. & Wiener, M. (2018). *R package randomForest: Breiman and Cutler's random forests for classification and regression* (Version 4.6-14). [Computer software]. The R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/randomForest/index.html>
- Linebaugh, C. W., & Lehner, L. H. (1977). Cueing hierarchies and word retrieval: A therapy program. In: Brookshire, R. H. (Ed.), *Clinical aphasiology: Vol. 7* (pp. 19-31). Minneapolis, MN: BRK Publishers.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.
- Lorenz, A. & Nickels, L. (2007). Orthographic cueing in anomic aphasia: How does it work? *Aphasiology*, 21(6-8), 670-686. <https://doi.org/10.1080/02687030701192182>
- Macoir, J., Routhier, S., Simard, A., & Picard, J. (2012). Maintenance and generalization effects of semantic and phonological treatments of anomia: A case study. *Communication Disorders Quarterly*, 33(2), 119-128. <https://doi.org/10.1177/1525740110387412>

- Marcotte, K., Adrover-Roig, D., Damien, B., de Préaumont, M., Génereux, S., Hubert, M., & Ansaldo, A. I. (2012). Therapy-induced neuroplasticity in chronic aphasia. *Neuropsychologia*, 50(8), 1776-1786.
<https://doi.org/10.1016/j.neuropsychologia.2012.04.001>
- Martin, N. (2013). Disorders of word production. In *Aphasia and related neurogenic communication disorders*. Papathanasiou, I., Coppens, P., & Potagas, C. (Eds). Burlington, MA: Jones & Barlett Learning.
- Mason, C., Nickels, L., McDonald, B., Moses, M., Makin, K. & Taylor, C. (2011). Treatment of word retrieval impairments in aphasia: Evaluation of a self-administered home programme using personally chosen words. *Aphasiology*, 25(2), 245-268.
<https://doi.org/10.1080/02687038.2010.489258>
- McNeil, M. R., Odell, K., & Tseng, C-H. (1991). Toward the integration of resource allocation into a general theory of aphasia. In T. E. Prescott (Ed.). *Clinical Aphasiology*, Vol. 20 (pp. 21-39). Austin, TX: Pro-Ed.
- Measso, G., Zappalà, G., Cavarzeran, F., Crook, T. H., Romani, L., Pirozzolo, F. J., Grigoletto, F., Amaducci, L. A., Massari, D., & Lebowitz, B. D. (1993). Raven's colored progressive matrices: A normative study of a random sample of healthy adults. *Acta Neurologica Scandinavica*, 88(1), 70-74. <https://doi.org/10.1111/j.1600-0404.1993.tb04190.x>
- Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., Thomas, M., Kramer, K., Lohmann, H., Flöel, A., Knecht, S., & Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience*, 10, 118.
<https://doi.org/10.1186/1471-2202-10-118>
- Miceli, G., Armitrano, A., Capasso, R., & Caramazza, A. (1996). The treatment of anomia resulting from output lexical damage: Analysis of two cases. *Brain and Language*, 52(1), 150-174. <https://doi.org/10.1006/brln.1996.0008>
- Minkina, I., Oelke, M., Bislick, L. P., Brookshire, C. E., Pompon, R. H., Silkes, J. P., & Kendall, D. L. (2016). An investigation of aphasic naming error evolution following phonomotor treatment. *Aphasiology*, 30(8), 962-980.
<https://doi.org/10.1080/02687038.2015.1081139>

- Minkina, I., Silkes, J. P., Bislick, L., Madden, E. B., Lai, V., Pompon, R. H., Torrence, J., Zimmerman, R. M., & Kendall, D. L. (2019). The influence of phonomotor treatment on word retrieval: Insights from naming errors. *Journal of Speech, Language, and Hearing Research*, 62(11), 4080-4104. https://doi.org/10.1044/2019_JSLHR-L-19-0014
- Mitchell, C., Gittins, M., Tyson, S., Vail, A., Conroy, P., Paley, L., & Bowen, A. (2021). Prevalence of aphasia and dysarthria among inpatient stroke survivors: Describing the population, therapy provision and outcomes on discharge. *Aphasiology*, 35(7), 950-960. <https://doi.org/10.1080/02687038.2020.1759772>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *BMJ*, 339: b2535. <https://doi.org/10.1136/bmj.b2535>
- Monsell, S., Matthews, G. H., & Miller, D. C. (1992). Repetition of lexicalization across languages: A further test of the locus of priming. *The Quarterly Journal of Experimental Psychology Section A*, 44(4), 763-783. <https://doi.org/10.1080/14640749208401308>
- Morgan, J. N. & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302), 415-434. <https://doi.org/10.2307/2283276>
- Morley, S. (2018). *Single case methods in clinical psychology*. Abingdon, Oxon: Routledge.
- Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P.A. Kolars, M. E. Wrolstad, & H. Bouma (Eds.). *NATO conference series: Vol. 13. Processing of Visible Language 2* (pp. 259-268). New York, N.Y.: Plenum Press. https://doi.org/10.1007/978-1-4684-0994-9_15
- Moss Rehabilitation Research Institute. (1996). *Detailed guide to Philadelphia Naming Test (PNT) scoring*. https://mrri.org/wp-content/uploads/2016/01/PNT-Glossary-of-Terms_scoring-guide.pdf
- Murdoch, B. E. (1990). *Acquired speech and language disorders: A neuroanatomical and functional neurological approach*. London, UK: Chapman and Hall.

- Murray, L. L. (2012). Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech-Language Pathology*, 21(2), S51-S64. [https://doi.org/10.1044/1058-0360\(2012/11-0067\)](https://doi.org/10.1044/1058-0360(2012/11-0067))
- Nakamura, K., Kuo, W-J., Pegado, F., Cohen, L., Tzeng, O. J. L., & Dehaene, S. (2012). Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proceedings of the National Academy of Sciences*, 109(50), 20762-20767. <https://doi.org/10.1073/pnas.1217749109>
- Nickels, L. (2002a). Improving word finding: Practice makes (closer to) perfect? *Aphasiology*, 16(10-11), 1047-1060. <https://doi.org/10.1080/02687040143000618>
- Nickels, L. (2002b). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, 16(10-11), 935-979. <https://doi.org/10.1080/02687030244000563>
- Nickels, L. & Best, W. (1996a). Therapy for naming disorders (Part I): Principles, puzzles and progress. *Aphasiology*, 10(1), 21-47. <https://doi.org/10.1080/02687039608248397>
- Nickels, L. & Best, W. (1996b). Therapy for naming disorders (Part II): Specifics, surprises and suggestions. *Aphasiology*, 10(2), 109-136. <https://doi.org/10.1080/02687039608248401>
- Nickels, L. & Cole-Virtue, J. (2004). Reading tasks from PALPA: How do controls perform on visual lexical decision, homophony, rhyme, and synonym judgements? *Aphasiology*, 18(2), 103-126. <https://doi.org/10.1080/02687030344000517-633>
- Nickels, L. & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, 33(10), 1281-1303. [https://doi.org/10.1016/0028-3932\(95\)00102-9](https://doi.org/10.1016/0028-3932(95)00102-9)
- Nightingale, A. M. (1993). *Normative data for 6 sub-tests of the PALPA: Evidence from the normal elderly population* [Unpublished thesis]. University College London.
- Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, 143(9), 992-1009. <https://doi.org/10.1037/bul0000108>

- Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A. (2016). The impact of dose on naming accuracy with persons with aphasia. *Aphasiology*, 30(9), 983-1011.
<https://doi.org/10.1080/02687038.2015.1100705>
- Olsen, E., Freed, D. B., & Marshall, R. C. (2012). Generalisation of personalised cueing to enhance word finding in natural settings. *Aphasiology*, 26(5), 618-631.
<https://doi.org/10.1080/02687038.2012.676851>
- Paciaroni, M. & Acciarresi, M. (2019). Poststroke fatigue. *Stroke*, 50(7), 1927-1933.
<https://doi.org/10.1161/STROKEAHA.119.023552>
- Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology*, 13(6), 503–511. [https://doi.org/10.1016/S0887-6177\(97\)00037-1](https://doi.org/10.1016/S0887-6177(97)00037-1)
- Patterson, J. P. (2001). The effectiveness of cueing hierarchies as a treatment for word retrieval impairment. *Neurophysiology and Neurogenic Speech and Language Disorders*, 11(2), 11-18. <https://doi.org/10.1044/nnsld11.2.11>
- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>
- Predovan, D., Prime, D., Arguin, M., Gosselin, F., Dell'Acqua, R., & Jolicœur, P. (2009). On the representation of words and nonwords in visual short-term memory: Evidence from human electrophysiology. *Psychophysiology*, 46(1):191-199.
<https://doi.org/10.1111/j.1469-8986.2008.00753.x>
- Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke*, 32(7), 1621-1626. <https://doi.org/10.1161/01.STR.32.7.1621>
- R Core Team (2019). *R: A language and environment for statistical computing* (Version 3.6.1) [Computer software]. The R Foundation for Statistical Computing. <https://www.r-project.org>

- Rapp, B. & Goldrick, M. (2006). Speaking words: Contributions of cognitive neuropsychological research. *Cognitive Neuropsychology*, 23(1), 39-73.
<https://doi.org/10.1080/02643290542000049>
- Raven, J. C. (1976). *Colored Progressive Matrices*. San Antonio, TX: Pearson.
- Raven, J. & Raven, J. (2003). Raven Progressive Matrices. In McCallum, R. S. (Ed.), *Handbook of nonverbal assessment* (pp. 223-237). Springer Science + Business Media New York. https://doi.org/10.1007/978-1-4615-0153-4_11
- Raymer, A. M. & Roitsch, J. (2023). Word retrieval treatments in aphasia: A survey of professional practice. *Aphasiology*, 37(7), 954-979.
<https://doi.org/10.1080/02687038.2022.2063791>
- REhabilitation and recovery of peopLE with Aphasia after StrokeE (RELEASE) Collaborators. Dosage, intensity, and frequency of language therapy for aphasia: A systematic review-based, individual participant data network meta-analysis. *Stroke*, 53(3), 956-967.
<https://doi.org/10.1161/STROKEAHA.121.035216>
- Renvall, K. & Nickels, L. (2019). Using treatment to improve the production of emotive adjectives in aphasia: A single-case study. *Aphasiology*, 33(11), 1348-1371.
<https://doi.org/10.1080/02687038.2019.1643001>
- Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-Language Pathology*, 17(2), 161-172. <https://doi.org/1058-0360/08/1702-0161>
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–133.
- Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Cambridge, U. K.: Pearson.
- Rose, M., Copland, D., Nickels, L., Togher, L., Godecke, E., Rai, T., Meinzer, M., Cadilhac, D., Kim, J., Hurley, M., Wilcox C., Steel, G, Pierce, J, & Carragher, M. (2021, September 10). *Main results and predictors of treatment response in the COMPARE randomized controlled trial of Constraint Induced Aphasia Therapy-Plus, Multimodality Aphasia*

- Therapy and Usual Care for chronic post-stroke aphasia* [Paper presentation]. British Aphasiology Society International Biennial Conference 2021 (Virtual), Newcastle, United Kingdom.
- Rose, M. & Douglas, J. (2001). The differential facilitatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology*, 15(10-11), 977-990.
<https://doi.org/10.1080/02687040143000339>
- Rose, M. & Douglas, J. (2008). Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology*, 22(1), 20-41.
<https://doi.org/10.1080/02687030600742020>
- Routhier, S., Bier, N., & Macoir, J. (2015). The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia. *Journal of Communication Disorders*, 54, 43-55. <https://doi.org/10.1016/j.jcomdis.2015.01.003>
- Routhier, S., Bier, N., & Macoir, J. (2016). Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia. *Aphasiology*, 30(2-3), 269-289.
<https://doi.org/10.1080/02687038.2014.973361>
- Russo, M. J., Prodan, V., Meda, N. N., Carcavallo, L., Muracioli, A., Sabe, L., Bonamico, L., Allegri, R. F., & Olmos, L. (2017). High-technology augmentative communication for adults with post-stroke aphasia: A systematic review. *Expert Review of Medical Devices*, 14(5), 355-370. <https://doi.org/10.1080/17434440.2017.1324291>
- Sage, K., Snell, C., & Lambon Ralph, M. A. (2011). How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation*, 21(1), 26-41.
<https://doi.org/10.1080/09602011.2010.528966>
- Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, 520, 609-611. <https://doi.org/10.1038/520609a>
- Schwartz, M. F., Middleton, E. L., Brecher, A., Gagliardi, M., & Garvey, K. (2016). Does naming accuracy improve through self-monitoring of errors? *Neuropsychologia*, 84(April), 272-281. <https://doi.org/10.1016/j.neuropsychologia.2016.01.027>

- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568.
<https://doi.org/10.1037/0033-295X.96.4.523>
- Siegel, S. & Castellan, Jr., N. John (1988). *Nonparametric statistics for the behavioural sciences* (2nd ed.). US: McGraw-Hill, Inc.
- Shah, A. D., Barlett, J. D., Carpenter, J., Nicolas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*, 179(6), 764-774.
<https://doi.org/10.1093/aje/kwt312>
- Silkes, J. P., Dierkes, K. E., & Kendall, D. L. (2013) Masked repetition priming effects on naming in aphasia: A phase I treatment study. *Aphasiology*, 27(4), 381-397.
<https://doi.org/10.1080/02687038.2012.745475>
- Smirni, P., Villardita, C., & Zappalà, G. (1983). Influence of different paths on spatial memory performance in the block-tapping test. *Journal of Clinical and Experimental Neuropsychology*, 5(4), 355-359. <https://doi.org/10.1080/01688638308401184>
- Snell, C., Sage, K., & Lambon Ralph, M. A. (2010). How many words should we provide in anomia therapy? A meta-analysis and a case series study. *Aphasiology*, 24(9), 1064-1094. <https://doi.org/10.1080/02687030903372632>
- Steinberg, D. (n. d.). Rules of thumb when working with small data samples.
<https://www.salford-systems.com/blog/dan-steinberg/rules-of-thumb-when-working-with-small-data-samples>
- Stimley, M. A. & Noll, J. D. (1991). The effects of semantic and phonemic prestimulation cues on picture naming in aphasia. *Brain and Language*, 41(4):496-509.
[https://doi.org/10.1016/0093-934x\(91\)90170-6](https://doi.org/10.1016/0093-934x(91)90170-6)
- Strobl, C., Boulesteix, A-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8, 25.
<https://doi.org/10.1186/1471-2105-8-25>

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! A new, conditional variable importance measure for random forests available in the party package. *The R Journal*, 2, 14-17.

Stroke Association (n.d.). What is aphasia? <https://www.stroke.org.uk/what-is-aphasia>

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. Hove, East Sussex: Psychology Press.

Sze, W. P., Hameau, S., Warren, J., & Best, W. (2021). Identifying the components of a successful spoken naming therapy: a meta-analysis of word-finding interventions for adults with aphasia. *Aphasiology*, 35(1), 33-72.
<https://doi.org/10.1080/02687038.2020.1781419>

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavioral Research Methods*, 46(1), 263-273. <https://doi.org/10.3758/s13428-013-0355-9>

Tactus Therapy Solutions Ltd. (2022). Naming therapy: Word finding for aphasia [Mobile app]. App store. <https://apps.apple.com/app/naming-therappy/id451093640>

Tagliamonte, S. A. & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135-178. <https://doi.org/10.1017/S0954394512000129>

Tesak, J. & Code, C. (2008). *Milestones in the history of aphasia: Theories and protagonists*. Hove, East Sussex: Psychology Press.

The GIMP Team (2019). GIMP (Version 2.10.6) [Software]. <https://www.gimp.org/>

The importance of no evidence. [Editorial]. (2019, March). *Nature Human Behaviour*, March 2019 (3), 197. <https://doi.org/10.1038/s41562-019-0569-7>

Tyler, C., Finch, E., Shrubsole, K., Ryan, B., Soroli, E., Martinez-Ferreiro, S., & Wallace, S. J. (2023). Aphasia outcome measurement in clinical practice: An international survey. *Aphasiology*, 37(10), 1576-1593. <https://doi.org/10.1080/02687038.2022.2112143>

- Urbaniak, G. C., & Plous, S. (2013). Research Randomizer (Version 4.0) [Computer software]. <http://www.randomizer.org/>
- van Hees, S., Angwin, A., McMahon, K., & Copland, D. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia, *Neuropsychological Rehabilitation*, 23(1), 102-132. <https://doi.org/10.1080/09602011.2012.726201>
- van Scherpenberg, C., Fieder, N., Savage, S., & Nickels, L. (2019). The relationship between response consistency in picture naming and storage impairment in people with semantic variant primary progressive aphasia. *Neuropsychology*, 33(1), 13-34. <https://doi.org/10.1037/neu0000485>
- Viggiano, M. P., Vannucci, M., & Righi, S. (2004). A new standardized set of ecological pictures for experimental and clinical research on visual object processing. *Cortex*, 40(3), 491-509. [https://doi.org/10.1016/S0010-9452\(08\)70142-4](https://doi.org/10.1016/S0010-9452(08)70142-4)
- Vitali, P., Tettamanti, M., Abutalebi, J., Ansaldo, A.-I., Perani, D., Cappa, S. F., & Joanette, Y. (2010). Generalization of the effects of phonological training for anomia using structural equation modelling: A multiple single-case study. *Neurocase*, 16(2), 93-105. <https://doi.org/10.1080/13554790903329117>
- Walker, G. M., Basilakos, A., Fridriksson, J., & Hickok, G. (2022). Beyond percent correct: Measuring change in individual picture naming ability. *Journal of Speech, Language, & Hearing Research*, 65(1), 215-237. https://doi.org/10.1044/2021_JSLHR-20-00205
- Wallace, S. E. & Kimelman, M. D. Z. (2013). Generalization of word retrieval following semantic feature treatment. *NeuroRehabilitation*, 32(4), 899-913. <https://doi.org/10.3233/NRE-130914>
- Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A-C., Marshall, J., Nicholas, M., Patterson, J., Pearl, G., Rochon, E., Rose, M, Sage, K., Small, S., & Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, 14(2), 180-185. <https://doi.org/10.1177/1747493018806200>

- Wambaugh, J. L., Mauszycki, S., & Wright, S. (2014). Semantic feature analysis: Application to confrontation naming of actions in aphasia. *Aphasiology*, 28(1), 1-24.
<https://doi.org/10.1080/02687038.2013.845739>
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation & Developmental Disabilities Research Reviews*, 13(1), 70-77.
doi:10.1002/mrdd.20139
- Wechsler, D. & Stone, C. P. (1972). *Wechsler Memory Scale*. New York: N.Y.: The Psychological Corporation.
- Wei, C-L., Rowe, G. T., Escobar-Briones, E., Boetius, A., Soltwedel, T., Caley, M. J., Soliman, Y., Huettmann, F., Qu, F., Yu, Z., Pitcher, C. R., Haedrich, R. L., Wicksten, M. K., Rex, M. A., Baguley, J. G., Sharma, J., Danovaro, R., MacDonald, I. R., Nunnally, C. C., Deming, J. W., Montagna, P., Levesque, M., Weslawski, J. M., Wlodarska-Kowalczyk, M., Ingole, B. S., Bett, B. J., Billett, D. S. M., Yool, A., Bluhm, B. A., Iken, K., & Narayanaswamy, B. E. (2010). Global patterns and predictions of seafloor biomass using random forests. *PLoS ONE*, 5(12), e15323. <https://doi.org/10.1371/journal.pone.0015323>
- Weidner, K & Lowman, J. (2020). Telepractice for adult speech-language pathology services: A systematic review. *Perspectives of the ASHA Special Interest Groups: SIG 18*, 5(1), 326-338. https://doi.org/10.1044/2019_PERSP-19-00146
- Weill-Chounlamountry, A., Capelle, N., Tessier, C., & Pradat-Diehl, P. (2013). Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction. *Brain Injury*, 27(5), 620-631. <https://doi.org/10.3109/02699052.2013.767936>
- Whitworth, A., Webster, J., & Howard, D. (2005). *Cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide*. Hove, East Sussex: Psychology Press.
- Wilssens, I., Vandenborre, D., van Dun, K., Verhoeven, J., Visch- Brink, E., & Marien, P. (2015). Constraint-induced aphasia therapy versus intensive semantic treatment in fluent aphasia. *American Journal of Speech-Language Pathology*, 24(2): 281-294.
https://doi.org/10.1044/2015_AJSLP-14-0018

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

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

World Health Organization. (2020). *Transmission of SARS-CoV-2: Implications for infection prevention precautions* (Scientific brief, 09 July 2020). World Health Organization. <https://apps.who.int/iris/handle/10665/333114>

Wright, H. H. & Fergadiotis, G. (2012). Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology*, 26(3/4), 258-278. <https://doi.org/10.1080/02687038.2011.604304>

Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L. (2008). Using a written cueing hierarchy to improve verbal naming in aphasia. *Aphasiology*, 22(5), 522-536. <https://doi.org/10.1080/02687030701487905>

Appendix 2.1

List of information sources

(1) For electronic databases, we included the more general search engines (PsychINFO, Web of Science, Embase, PubMed, Cochrane Library Databases) as well as discipline-specific search engines (for allied health: Allied and Complementary Medicine (AMED), Cumulative Index to Nursing and Allied Health Literature (CINAHL Plus); for speech and language therapy: Speech Pathology Database for Best Interventions and Treatment Efficacy (SpeechBITE)).

(2) For trial registers, we included the following: ClinicalTrials.gov, the Stroke Trials Registry (<http://www.strokecenter.org/trials/>), Current Controlled Trials, International Standard Registered Clinical/soCial sTudy Number (ISRCTN) registry, World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), Database of Research in Stroke (DORIS) and the Cochrane Central Register of Controlled Trials.

(3) Lastly, grey literature was trawled using ProQuest for unpublished theses/dissertations and by hand-searching reference lists of other reviews for articles not identified electronically.

Appendix 2.2

Articles eligible for the meta-analysis

SN	Author(s)	Year	Title of article	Journal
1	Abel, S., Weiller, C., Huber, W., Wilmes, K., & Specht, K.	2015	Therapy-induced brain reorganization patterns in aphasia	Brain
2	Adrian, J. A., Gonzalez, M., Buiza, J. J., & Sage, K.	2011	Extending the use of Spanish Computer-assisted Anomia Rehabilitation Program (CARP-2) in people with aphasia	Journal of Communication Disorders
3	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	A comparison of word versus sentence cues as therapy for verb naming in aphasia	Aphasiology
4	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	Errorless and errorful therapy for verb and noun naming in aphasia	Aphasiology
5	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech?	International Journal of Language and

				Communication Disorders
6	Conroy, P., Sage, K., & Lambon Ralph, M. A.	2009	The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia	Aphasiology
7	Conroy, P. & Scowcroft, J.	2012	Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study	Neuropsychological Rehabilitation
8	Dechene, L., Tousignant, M., Boissy, P., Macoir, J., Heroux, S., Hamel, M., Briere, S., & Page, C.	2011	Simulated in-home teletreatment for anomia	International Journal of Telerehabilitation
9	DeLong, C., Nessler, C., Wright, S., & Wambaugh, J.	2015	Semantic feature analysis: Further examination of outcomes	American Journal of Speech-Language Pathology
10	Harnish, S. M., Morgan, J., Lundine, J. P., Bauer, A., Singletary, F., Benjamin, M. L., Gonzalez Rothi, L. J., & Crosson, B.	2014	Dosing of a cued picture-naming treatment for anomia	American Journal of Speech-Language Pathology

11	Hashimoto, N.	2016	The use of one or three semantic associative primes in treating anomia in aphasia	American Journal of Speech-Language Pathology
12	Hendricks, C. T., Nicholas, M. L., & Zipse, L.	2014	Effects of phonological neighbourhood on the treatment of naming in aphasia	Aphasiology
13	Jacquemot, C., Dupoux, E., Robotham, L., & Bachoud-Levi, A. C.	2012	Specificity in rehabilitation of word production: A meta-analysis and a case study	Behavioural Neurology
14	Kiran, S.	2008	Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity	Journal of Speech, Language, and Hearing Research
15	Kiran, S. & Johnson, L.	2008	Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories	American Journal of Speech-Language Pathology
16	Macoir, J., Routhier, S., Simard, A., & Picard, J.	2012	Maintenance and generalization effects of semantic and phonological treatments of anomia: A case study	Communication Disorders Quarterly

17	Marcotte, K., Adrover-Roig, D., Damien, D., de Preaumont, M., Genereux, S., Hubert, M., Ansaldo, A. I.	2012	Therapy-induced neuroplasticity in chronic aphasia	Neuropsychologia
18	Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., Thomas, M., Kramer, K., Lohmann, H., Flöel, A., Knecht, S., & Breitenstein, C.	2009	Imaging short- and long-term training success in chronic aphasia	BMC Neuroscience
19	Off, C. A., Griffin, J. R., Spencer, K. A., & Rogers, M. A.	2016	The impact of dose on naming accuracy with persons with aphasia	Aphasiology
20	Olsen, E., Freed, D. B., & Marshall, R. C.	2012	Generalisation of personalised cueing to enhance word finding in natural settings	Aphasiology
21	Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L.	2008	Using semantic feature analysis to improve contextual discourse in adults with aphasia	American Journal of Speech-Language Pathology
22	Rose, M. & Douglas, J.	2008	Treating a semantic word production deficit in aphasia with verbal and gesture methods	Aphasiology

23	Routhier, S., Bier, N., & Macoir, J.	2015	The contrast between cueing and/or observation in therapy for verb retrieval in post-stroke aphasia	Journal of Communication Disorders
24	Routhier, S., Bier, N., & Macoir, J.	2016	Smart tablet for smart self-administered treatment of verb anomia: Two single-case studies in aphasia	Aphasiology
25	Sage, K., Snell, C., & Lambon Ralph, M. A.	2011	How intensive does anomia therapy for people with aphasia need to be?	Neuropsychological Rehabilitation
26	Snell, C., Sage, K., & Lambon Ralph, M. A.	2010	How many words should we provide in anomia therapy? A meta-analysis and a case series study.	Aphasiology
27	van Hees, S., Angwin, A., McMahon, K., & Copland, D.	2013	A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia	Neuropsychological Rehabilitation
28	Vitali, P., Tettamanti, M., Abutalebi, J., Ansaldo, A.-I., Perani, D., Cappa, S. F., & Joanette, Y.	2010	Generalization of the effects of phonological training for anomia using structural equation modelling: A multiple single-case study.	Neurocase
29	Wallace, S. E. & Kimelman, M. D. Z.	2013	Generalization of word retrieval following semantic feature treatment	NeuroRehabilitation

30	Wambaugh, J. L., Mauszycki, S., & Wright, S.	2014	Semantic feature analysis: Application to confrontation naming of actions in aphasia	Aphasiology
31	Weill-Chounlamountry, A., Capelle, N., Tessier, C., & Pradat-Diehl, P.	2013	Multimodal therapy of word retrieval disorder due to phonological encoding dysfunction	Brain Injury
32	Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L.	2008	Using a written cueing hierarchy to improve verbal naming in aphasia	Aphasiology

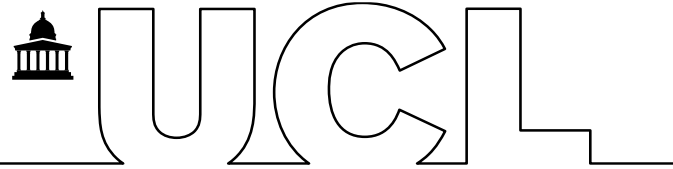
Appendix 2.3

List of therapy components based on the 'RITA' framework

Information about Regimen		(continued) Techniques	
No. of minutes per session		Copying of written target word: Compulsory	✓: X :
No. of sessions per week		Semantic cue(s) generation	✓: X :
No. of sessions for entire therapy		Phonological cue(s) generation	✓: X :
Information about Items		Semantic task(s)	✓: X :
Total no. of items treated across sessions		Orthographic task(s)	✓: X :
Total no. of times each item was named across sessions			
Grammatical class of treated words	N, V etc.		
Information about Techniques		Information about Technique Application	
Provision of semantic cue(s): Explicit	✓: X :	Application of cues: Progression	↑: ↓:
Provision of phonological cue(s): Explicit	✓: X :	Application of cues: Depends on response vs. compulsory vs. mixture	
Provision of orthographic cue(s): Explicit	✓: X :	Feedback on accuracy of the naming response	Explicit: Implicit: None:
Provision of phonological cue(s): Implicit	✓: X :	Termination criteria	
Provision of the target word in spoken form	✓: X :	Delivery mode for therapy: Computer; in-person; mixture	
Provision of the target word in written form	✓: X :	Homework	✓: X :
Spoken repetition of target word: Compulsory	✓: X :		
Spoken repetition of target word: Contingent on response	✓: X :		

LONDON'S GLOBAL UNIVERSITY

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES



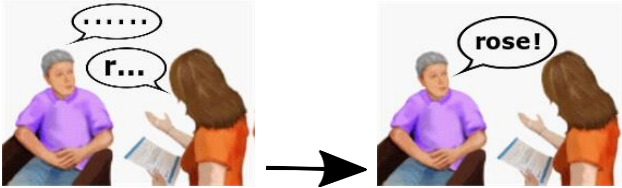
Call for Research Participants

- Do you know **anyone** who had a **stroke**?
- Do you know **anyone** who has **problems finding or saying** the right word?

If you answered **“YES”** to the **two** questions:

We would like to **invite you OR someone you know** to

take part in a research project.

		In this research we will find out more about naming .
		We will research the type of cues that can help people find words.

To find out more, please **contact** the researcher at:



[Email address inserted in the actual poster]



[Mobile number inserted in the actual poster]

Appendix 3.2

Brief description of baseline tasks administered in the facilitation study

Name of task/test	Description
Oromotor speech screeners	
Diadochokinetic task	• Rapid repetition of /pataka/ for a minute. • Count the number of intelligible /pataka/ produced. • Norms followed Icht and Ben-David (2014): “Lower boundary for non-pathological performance of 4.6 syllables/s” (p. 34)
Oromotor task	To assess the oro-masculature integrity and to rule out orofacial apraxia. Instructions: • Open your mouth (1-step) • Say ‘ah’ • Stick out your tongue (1-step) • Lick your lips (1-step): Check for lingual control • Kiss a baby (1-step): Check for lip rounding • Give me a smile (1-step): Check for lip spread • Show me your teeth (1-step) • Close your lips tightly, and do not let me open: Check for lip seal
Spoken language processing	
PALPA 23. Spoken letter sound-written letter matching	• Choose the correct printed alphabet (choice of four) that corresponds to the spoken phoneme given by the tester: e.g., /b/ for ‘b’
PALPA 23. Spoken letter name-written letter matching	• Choose the correct printed alphabet (choice of four) that corresponds to the spoken letter name given by the tester: e.g., /bi/ for ‘b’
PALPA 5. Auditory lexical decision task	• Indicate ‘Yes’ for spoken real words given by the tester, and ‘No’ to the spoken pseudo-words
CAT 14. Repetition of nonwords	• Verbally repeat the spoken pseudo-word given by the tester, e.g., ‘trimpy’
PALPA 8. Nonword repetition task	• Verbally repeat the spoken pseudo-word given by the tester, e.g., ‘vater’

Name of task/test	Description
Written language processing	
PALPA 18. Letter discrimination: Mirror reversal	Indicate 'Yes' for letters printed in the correct orientation
PALPA 22. Letter sounding	Produce the sound corresponding to the printed alphabet, e.g., /p/ for 'p'
PALPA 22. Letter naming	Say out the name of the printed alphabet: e.g., /pi/ for 'p'
PALPA 25. Visual lexical decision task	Indicate 'Yes' for real words printed on the sheet, and 'No' to the pseudo-words
CAT 20. Reading words	Read aloud the words (single morphemes, e.g., 'family') printed on the sheet
CAT 21. Reading complex words	Read aloud the words (multi-morphemic words, e.g., 'recooked') printed on the sheet
CAT 22. Reading function words	Read aloud the words (e.g., determiners like 'but') printed on the sheet
CAT 23. Reading nonwords	Read aloud the pseudo-words printed on the sheet (e.g., 'fask')
PALPA 36. Reading nonwords	Read aloud the pseudo-words printed on the sheet (e.g., 'grest')
Semantic processing	
CAT 2. Semantic memory	- Similar to Pyramid and Palm Trees: Three pictures version - Choose the picture (choice of four) that is related to the target picture in the middle, e.g., 'banana' with target picture, 'monkey'
CAT 7. Comprehension of spoken words	Choose the correct picture (choice of four) that corresponds to the spoken target word given by the tester
CAT 8. Comprehension of written words	Choose the correct picture (choice of four) that corresponds to the written target word presented in the middle
PALPA 49. Auditory synonym judgement	Indicate 'Yes' if the two spoken words are close in meaning. Indicate 'No' if otherwise.
PALPA 50. Written synonym judgement	Indicate 'Yes' if the two written words are close in meaning. Indicate 'No' if otherwise.
Naming	
CAT 17. Naming objects	Name the target item represented in the picture (e.g., 'car')
CAT 19. Naming actions	Name the target action represented in the picture (e.g., 'eat'/'eating')
CAT 25. Writing picture names	Write out the target word for the picture (all stimuli are nouns; e.g., 'hat')

Appendix 3.3

Sources of photographs used as stimuli ($n = 296$) in the facilitation study (Chapters 3 and 4)

Source	Number of photographs
Adlington, Laws, and Gale (2009)	11 (3.72%)
Brodeur, Dionne-Dostie, Montreuil, and Lepage (2010)	60 (20.27%)
Creet, Morris, Howard, and Nickels (2019)	167 (56.42%)
Viggiano, Vannucci, and Righi (2004)	32 (10.81%)
Google search for copyrights-free photographs (e.g., www.stickpng.com ; https://pixabay.com)	26 (8.78%)

Appendix 3.4

Changes in the average number of items in raw count correctly named and improvement in average RT (from baseline immediately prior to therapy) (SD) across participants in the facilitation study. Only post hoc results that remained significant after comparison against repeated exposure data are highlighted.

Time Points	Difference between post-facilitation session 1 (15 minutes later) and baseline 2			Difference between post-facilitation session 2 (1 week later) and baseline 2		
	Mean initial naming accuracy	Mean final naming accuracy	RT (ms) ¹	Mean initial naming accuracy	Mean final naming accuracy	RT (ms) ¹
Initial Letter: Orthography	3.20 (3.27) [†] (Propn Δ: 0.16 (0.17))	3.60 (4.51) (Propn Δ: 0.19 (0.21))	1349.41 (1424.35) * (Propn Δ: 0.15 (0.15))	2.80 (1.30) * (Propn Δ: 0.13 (0.10))	3.00 (1.00) * (Propn Δ: 0.16 (0.093))	671 (342.02) * (Propn Δ: 0.073 (0.045))
Whole Word: Orthography	3.60 (1.95) * (Propn Δ: 0.16 (0.080))	6.80 (2.95) * (Propn Δ: 0.35 (0.26))	2327.11 (1766.28) * (Propn Δ: 0.24 (0.19))	4.20 (3.03) * (Propn Δ: 0.21 (0.22))	6.40 (2.61) * (Propn Δ: 0.33 (0.23))	2210.74 (1769.87) * (Propn Δ: 0.23 (0.19))
Initial Letter: Orthography and Phonology	1.00 (1.87) (Propn Δ: 0.054 (0.085))	3.20 (3.77) (Propn Δ: 0.19 (0.24))	1238.77 (1608.41) (Propn Δ: 0.13 (0.16))	1.40 (2.30) (Propn Δ: 0.083 (0.15))	3.00 (3.08) [†] (Propn Δ: 0.18 (0.25))	1069.13 (2039.19) (Propn Δ: 0.11 (0.20))
Whole Word: Orthography and Phonology	2.40 (2.61) (Propn Δ: 0.11 (0.10))	3.20 (4.97) (Propn Δ: 0.12 (0.23))	899.78 (1236.42) (Propn Δ: 0.091 (0.13))	1.80 (1.10) [†] (Propn Δ: 0.087 (0.058))	2.80 (3.63) (Propn Δ: 0.13 (0.14))	901.68 (693.88) (Propn Δ: 0.100 (0.073))
Repeated Exposure	1.40 (0.55) (Propn Δ: 0.060 (0.017))	3.80 (1.79) * (Propn Δ: 0.20 (0.16))	1103.45 (993.82) * (Propn Δ: 0.11 (0.11))	0.80 (1.30) (Propn Δ: 0.041 (0.072))	2.60 (2.79) (Propn Δ: 0.14 (0.19))	870.84 (1035.60) * (Propn Δ: 0.089 (0.11))
Unseen Control: Untreated, unexposed	-1.20 (2.05) (Propn Δ: -0.059 (0.082))	-0.20 (2.05) (Propn Δ: 0.014 (0.12))	-119.90 (664.01) (Propn Δ: -0.0078 (0.074))	1.00 (2.55) (Propn Δ: 0.056 (0.13))	2.40 (3.44) (Propn Δ: 0.10 (0.13))	605.79 (1121.63) (Propn Δ: 0.065 (0.11))

¹ To yield a positive number, for RTs, we subtract the post-therapy result from the baseline's.

² Propn = Proportional

Results based post hoc Wilcoxon signed ranks test: [†] .05 < *p* < .07, * .01 < *p* < .05, ** .001 < *p* < .01, *** *p* < .001

Appendix 3.5

Table summarizing the post-facilitation changes in terms of values for the studies involving the use of orthographic form as facilitation. Studies highlighted in yellow refer to those whose cue condition closely resembles a facilitation cue condition in Chapters 3 and 4.

	Group Level	Individual Level	
	Group Average	Largest Amount of Δ in Value	Smallest Amount of Δ in Value
Study 1:	Bruce and Howard (1988): $n = 11$ Facilitation technique: Orthographic task – A letter board with 9 letter choices (written)		
Post-facilitation duration:	5 seconds later		
Δ in Proportion:	0.13 ($SD = 0.128$)	0.47 (1 participant: PAB)	0 (3 participants: OLI, MIL, LIL)
Δ in Raw Values:	2.93 ($SD = 2.19$)	8	0
Study 2a:	Howard and Harding (1998): $n = 1$ (Participant SD) Facilitation technique: Orthographic task – A letter board with letter choices (written), number of choices not stated		
Post-facilitation duration:	5 seconds later		
Δ in Proportion:	N.A.	0.87	N.A.
Δ in Raw Values:	N.A.	20	N.A.
Study 2b:	Howard and Harding (1998): $n = 1$ (Participant SD) Facilitation technique: Orthographic cueing – Written ‘initial letter’ cue provided if the participant failed to name the picture		
Post-facilitation duration:	5 seconds later		
Δ in Proportion:	N.A.	0.96	N.A.
Δ in Raw Values:	N.A.	22	N.A.

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to the facilitation session

N.A.: Not relevant to paper

(continued)

	Group Level	Individual Level	
	Group Average	Largest Amount of Δ in Value	Smallest Amount of Δ in Value
Study 3:	Avila et al. (2001): $n = 1$ (Participant FR) Facilitation technique: Pre-stimulation with written word		
Post-facilitation duration:	5 seconds later		
Δ in Proportion:	N.A.	0.41	N.A.
Δ in Raw Values:	N.A.	36	N.A.
Study 4a:	Best et al. (2002): $n = 11$ Facilitation technique: Orthographic cueing – A single written CV cue provided if the participant failed to name the picture correctly within 5 seconds		
Post-facilitation duration:	5 seconds and 10 minutes later		
Δ in Proportion:	5 seconds post-facilitation: 0.47 10 mins post-facilitation: 0.41	Information not available in paper	Information not available in paper
Δ in Raw Values:	5 seconds post-facilitation: 5.62 10 mins post-facilitation: 4.98	Information not available in paper	Information not available in paper
Study 4b:	Best et al. (2002): $n = 11$ Facilitation technique: Orthographic task – A choice of two written CV cues presented if the participant failed to name the picture correctly within 5 seconds		
Post-facilitation duration:	5 seconds and 10 minutes later		
Δ in Proportion:	5 seconds post-facilitation: 0.42 10 mins post-facilitation: 0.39	Information not available in paper	Information not available in paper
Δ in Raw Values:	5 seconds post-facilitation: 4.99 10 mins post-facilitation: 4.65	Information not available in paper	Information not available in paper

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to the facilitation session

N.A.: Not relevant to paper

(continued)

	Group Level	Individual Level	
	Group Average	Largest Amount of Δ in Value	Smallest Amount of Δ in Value
Study 5:	Lorenz and Nickels (2007): $n = 3$		
	Facilitation technique: Pre-stimulation with a written initial letter		
Post-facilitation duration:	Directly after cue was administered and 15-20 minutes post-facilitation		
Δ in Proportion:	Directly after cue: 0.197 ($SD = 0.167$) 15-20 minutes post-facilitation: 0.233 ($SD = 0.209$)	Directly after cue: 0.39 (Participant KCC, Phase 1) 15-20 minutes post-facilitation: 0.57 (Participant KCC, Phase 1)	Directly after cue: 0.021 (Participant JUE, Phase 1) 15-20 minutes post-facilitation: 0.071 (Participant KCC, Phase 2)
Δ in Raw Values:	Directly after cue: 5.60 ($SD = 4.56$) 15-20 minutes post-facilitation: 6.80 ($SD = 5.63$)	Directly after cue: 11 (Participant KCC, Phase 1) 15-20 minutes post-facilitation: 16 (Participant KCC, Phase 1)	Directly after cue: 1 (Participant JUE, Phase 1) 15-20 minutes post-facilitation: 2 (Participant KCC, Phase 2)

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to the facilitation session

Appendix 4.1

Individual results based on WEST-Trend and ROC for each main naming outcome measure

P1:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 1.439246$, $p = .0803941042129381$, $d = 0.262769136406122$	<i>n.s.</i> $t(29) = -0.4788$, $p = .682164154569619$, $d = -0.0874164385489952$
Whole Word: Written Only	$\dagger$ $t(29) = 1.558899$, $p = .0649338867975472$, $d = 0.284614803887954$	<i>n.s.</i> $t(29) = 0.578682$, $p = .283635670613826$, $d = 0.105652414267222$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.947052$, $p = .175718795245577$, $d = 0.172907221412221$	<i>n.s.</i> $t(29) = 1.033985$, $p = .154844677499928$, $d = 0.188778920430145$
Whole Word: Written and Spoken	$\dagger$ $t(29) = 1.60591$, $p = .0595647384061346$, $d = 0.293197735804187$	<i>n.s.</i> $t(29) = 1.209208$, $p = .118174581863783$, $d = 0.220770233044039$
Repeated Exposure	<i>n.s.</i> $t(29) = 0.420511$, $p = .338606096724625$, $d = 0.076774402915877$	<i>n.s.</i> $t(29) = 0.972446$, $p = .169435155467599$, $d = 0.177543625992134$
Control: Untreated Unexposed	<i>n.s.</i> $t(29) = -0.17668$, $p = .569506653229152$, $d = -0.0322574857871394$	<i>n.s.</i> $t(29) = -0.68209$, $p = .749701552925421$, $d = -0.124531206492085$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	$*$ $t(29) = 2.307539$, $p = .0141789407131551$, $d = 0.421296983379124$	<i>n.s.</i> $t(29) = 0.682086$, $p = .250298447074579$, $d = 0.124531206492085$
Whole Word: Written Only	$***$ $t(29) = 3.672652$, $p = .000482674851121559$, $d = 0.670531433342819$	<i>n.s.</i> $t(29) = 1.186494$, $p = .122528533213292$, $d = 0.216623097436635$
Initial Letter and Phoneme: Written and Spoken	$*$ $t(29) = 1.988604$, $p = .0281274877487832$, $d = 0.363067737231197$	<i>n.s.</i> $t(29) = 0.471857$, $p = .320280103512398$, $d = 0.086148854200009$
Whole Word: Written and Spoken	$***$ $t(29) = 3.844095$, $p = .000304942747461234$, $d = 0.701832602636228$	$*$ $t(29) = 2.088277$, $p = .0228317920416613$, $d = 0.381265535449119$
Repeated Exposure	<i>n.s.</i> $t(29) = 1.354785$, $p = .0929735259175889$, $d = 0.247348811598431$	<i>n.s.</i> $t(29) = 1.262302$, $p = .108448701655301$, $d = 0.230463778311372$

Control: Untreated Unexposed	* $t(29) = 2.065116$, $p = .0239767203356889$, $d = 0.377036951431949$	<i>n.s.</i> $t(29) = -1.35479$, $p = .907026474082411$, $d = -0.247348811598431$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	* $t(29) = 1.805719$, $p = .0406765997299318$, $d = 0.329677728827308$	<i>n.s.</i> $t(29) = 0.527644$, $p = .300880446779293$, $d = 0.096334086890919$
Whole Word: Written Only	** $t(29) = 2.749664$, $p = .00508168202073321$, $d = 0.502017674382642$	<i>n.s.</i> $t(29) = 1.316189$, $p = .0992098132458178$, $d = 0.24030217813724$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 1.410352$, $p = .0845362651534139$, $d = 0.257493790670375$	<i>n.s.</i> $t(29) = 1.08372$, $p = .14370861080412$, $d = 0.197859279003979$
Whole Word: Written and Spoken	** $t(29) = 3.293511$, $p = .00130512610261312$, $d = 0.601310133109802$	† $t(29) = 1.654445$, $p = .0544110542798999$, $d = 0.302059031985949$
Repeated Exposure	<i>n.s.</i> $t(29) = 1.475662$, $p = .0754053387831953$, $d = 0.269417815093434$	<i>n.s.</i> $t(29) = 1.383458$, $p = .088541289081614$, $d = 0.252583726071555$
Control: Untreated Unexposed	* $t(29) = 1.969777$, $p = .0292407483834694$, $d = 0.359630376560639$	<i>n.s.</i> $t(29) = -1.18298$, $p = .876786701408316$, $d = -0.215980826109731$

† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

P2:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(15) = 2.766417$, $p = .00720119452285694$, $d = 0.69160416896561$	* $t(15) = 2.178723$, $p = .0228559974667346$, $d = 0.544680837913244$
Whole Word: Written Only	** $t(15) = 3.585686$, $p = .00135194111968876$, $d = 0.896421457000795$	<i>n.s.</i> $t(15) = 0.415227$, $p = .341926969666758$, $d = 0.103806849817175$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(15) = 1.414214$, $p = .0888596965510698$, $d = 0.353553390593274$	<i>n.s.</i> $t(15) = 0.57735$, $p = .28613376303598$, $d = 0.144337567297406$
Whole Word: Written and Spoken	* $t(15) = 2.335497$, $p = .0169087714655386$, $d = 0.583874208121142$	<i>n.s.</i> $t(15) = -0.43574$, $p = .665386251352505$, $d = -0.108936167582649$
Repeated Exposure	<i>n.s.</i> $t(15) = 0.128318$, $p = .449801095127422$, $d = 0.0320794157253855$	<i>n.s.</i> $t(15) = 0.510015$, $p = .308731337830633$, $d = 0.127503825172134$
Control: Untreated Unexposed	* $t(15) = 1.884337$, $p = .0395243790968573$, $d = 0.471084164300895$	<i>n.s.</i> $t(15) = -2.23607$, $p = .979515522022082$, $d = -0.559016994374947$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(15) = 3.000000$, $p = .00448636873861166$, $d = 0.75$	<i>n.s.</i> $t(15) = 0.39736$, $p = .348349622040389$, $d = 0.0993399267798783$
Whole Word: Written Only	*** $t(15) = 4.938711$, $p = .000089190044665227$, $d = 1.23467766401201$	** $t(15) = 3.523896$, $p = .00153464468222876$, $d = 0.8809739062391$
Initial Letter and Phoneme: Written and Spoken	** $t(15) = 2.777778$, $p = .00703826534495621$, $d = 0.6944444444444444$	* $t(15) = 2.26995$, $p = .0191932412568855$, $d = 0.567487431343483$
Whole Word: Written and Spoken	** $t(15) = 3.721889$, $p = .00102266073130732$, $d = 0.930472293260036$	<i>n.s.</i> $t(15) = -1.46385$, $p = .918062193172173$, $d = -0.3659625273557$
Repeated Exposure	** $t(15) = 2.697304$, $p = .00827399686014926$, $d = 0.674326010815448$	† $t(15) = 1.671729$, $p = .0576513353358741$, $d = 0.417932171985154$
Control: Untreated Unexposed	* $t(15) = 2.3063$, $p = .0178928720368501$, $d = 0.576574930348822$	<i>n.s.</i> $t(15) = 0.124035$, $p = .451467313894418$, $d = 0.0310086836473021$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(15) = 2.856738$, $p = .00600087274475754$, $d = 0.714184515895687$	<i>n.s.</i> $t(15) = 1.422043$, $p = .0877369002597279$, $d = 0.355510834665807$
Whole Word: Written Only	*** $t(15) = 4.574375$, $p = .000182540640467709$, $d = 1.14359368175204$	** $t(15) = 3.3059$, $p = .00240023816268004$, $d = 0.826475052447947$
Initial Letter and Phoneme: Written and Spoken	** $t(15) = 2.799899$, $p = .00673123734548492$, $d = 0.699974701892868$	* $t(15) = 2.318509$, $p = .017474981982635$, $d = 0.57962722511052$
Whole Word: Written and Spoken	*** $t(15) = 3.800361$, $p = .00087097188015908$, $d = 0.950090182208326$	<i>n.s.</i> $t(15) = -1.30135$, $p = .893614196636906$, $d = -0.3253383636663$
Repeated Exposure	† $t(15) = 1.720306$, $p = .0529669362053187$, $d = 0.430076377246753$	† $t(15) = 1.743252$, $p = .0508721576043044$, $d = 0.435812943909263$
Control: Untreated Unexposed	* $t(15) = 2.099082$, $p = .0265734775396904$, $d = 0.524770529810114$	<i>n.s.</i> $t(15) = -0.18587$, $p = .572482530947782$, $d = -0.0464681360171765$

† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

P3:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(24) = 2.57307$, $p = .00834405262127827$, $d = 0.514614016722142$	** $t(24) = 2.771281$, $p = .00530640295776646$, $d = 0.554256258422041$
Whole Word: Written Only	* $t(24) = 1.899306$, $p = .0348041444481359$, $d = 0.379861130582753$	<i>n.s.</i> $t(24) = 1.364576$, $p = .0925225318418541$, $d = 0.272915295688405$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(24) = 1.114773$, $p = .137993442161709$, $d = 0.222954564573318$	<i>n.s.</i> $t(24) = 0.569495$, $p = .287156299994597$, $d = 0.1138989594903$
Whole Word: Written and Spoken	* $t(24) = 2.220544$, $p = .0180380502253926$, $d = 0.444108888002161$	<i>n.s.</i> $t(24) = 1.423303$, $p = .083759451516581$, $d = 0.28466059261768$
Repeated Exposure	<i>n.s.</i> $t(24) = 0.146124$, $p = .442521844818003$, $d = 0.029224861448609$	<i>n.s.</i> $t(24) = 0.592157$, $p = .27964185240467$, $d = 0.118431305092758$
Control: Untreated Unexposed	<i>n.s.</i> $t(24) = -1.10004$, $p = .858886312232221$, $d = -0.220007639286771$	<i>n.s.</i> $t(24) = -1.49635$, $p = .926200996060703$, $d = -0.299270249466504$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(24) = 2.700919$, $p = .00624017728572357$, $d = 0.540183821630266$	*** $t(24) = 3.781253$, $p = .000457023871539847$, $d = 0.756250549498658$
Whole Word: Written Only	*** $t(24) = 3.56857$, $p = .000777203997385975$, $d = 0.713714056959817$	* $t(24) = 1.869027$, $p = .036935556284463$, $d = 0.373805374788839$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(24) = 1.385641$, $p = .0892997155651775$, $d = 0.27712812921102$	* $t(24) = 2.34216$, $p = .0139003854514754$, $d = 0.468432035015296$
Whole Word: Written and Spoken	* $t(24) = 1.729052$, $p = .0483222452207057$, $d = 0.345810316628213$	† $t(24) = 1.585532$, $p = .0629674470450577$, $d = 0.317106393520824$
Repeated Exposure	<i>n.s.</i> $t(24) = 1.045436$, $p = .153118608458329$, $d = 0.209087235667219$	<i>n.s.</i> $t(24) = 1.389551$, $p = .0887113668860343$, $d = 0.277910133951951$
Control: Untreated Unexposed	<i>n.s.</i> $t(24) = 0.182069$, $p = .428528780604621$, $d = 0.0364138277420643$	<i>n.s.</i> $t(24) = -1.57181$, $p = .935456910273311$, $d = -0.31436209919735$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	** $t(24) = 2.676745$, $p = .00659524018759799$, $d = 0.53534899422778$	** $t(24) = 3.291097$, $p = .00153901567884041$, $d = 0.658219431605238$
Whole Word: Written Only	** $t(24) = 2.958799$, $p = .00342125922017978$, $d = 0.591759785840807$	† $t(24) = 1.674276$, $p = .0535304316304758$, $d = 0.334855132082746$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(24) = 0.667295$, $p = .25547480911783$, $d = 0.133458932734058$	* $t(24) = 1.872262$, $p = .0367025472717908$, $d = 0.374452363670075$
Whole Word: Written and Spoken	* $t(24) = 1.906679$, $p = .0343016418648291$, $d = 0.381335835829692$	<i>n.s.</i> $t(24) = 1.30121$, $p = .102770299034464$, $d = 0.260241933724852$
Repeated Exposure	<i>n.s.</i> $t(24) = 1.00589$, $p = .162251418367706$, $d = 0.201178091630343$	<i>n.s.</i> $t(24) = 0.829161$, $p = .20759199741622$, $d = 0.165832284802658$
Control: Untreated Unexposed	<i>n.s.</i> $t(24) = -0.94264$, $p = .822370741372276$, $d = -0.188527775728832$	<i>n.s.</i> $t(24) = -1.54804$, $p = .932650146383741$, $d = -0.309607663220514$

† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

P4:**Accuracy of Initial Responses**

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 1.278203$, $p = .105657076932652$, $d = 0.233366865464473$	<i>n.s.</i> $t(29) = 0.626013$, $p = .268101547575004$, $d = 0.114293757760707$
Whole Word: Written Only	<i>n.s.</i> $t(29) = 1.439246$, $p = .080394104212938$, $d = 0.262769136406122$	<i>n.s.</i> $t(29) = 0$, $p = .50$, $d = 0$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.775667$, $p = .222112669384838$, $d = 0.141616824687528$	<i>n.s.</i> $t(29) = -0.70833$, $p = .757802012927642$, $d = -0.129322607434048$
Whole Word: Written and Spoken	<i>n.s.</i> $t(29) = -0.21472$, $p = .584256011044523$, $d = -0.0392024354931636$	<i>n.s.</i> $t(29) = -0.42266$, $p = .662170938369094$, $d = -0.0771675155879229$
Repeated Exposure	<i>n.s.</i> $t(29) = 1.351581$, $p = .0934793535597681$, $d = 0.246763831327534$	<i>n.s.</i> $t(29) = 0.493657$, $p = .312633216814266$, $d = 0.0901290702255376$
Control: Untreated Unexposed	<i>n.s.</i> $t(29) = 0$, $p = .50$, $d = 0$	<i>n.s.</i> $t(29) = -1.83633$, $p = .961709562714244$, $d = -0.335265716671409$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 1.360538$, $p = .092070671145986$, $d = 0.248399129748022$	<i>n.s.</i> $t(29) = -0.3667$, $p = .641746482206886$, $d = -0.0669494261131333$
Whole Word: Written Only	<i>*</i> $t(29) = 1.997892$, $p = .0275920727749791$, $d = 0.36476353224648$	<i>n.s.</i> $t(29) = 0.24229$, $p = .405130389349751$, $d = 0.0442359030054757$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.862316$, $p = .197794209892833$, $d = 0.157436732646244$	<i>n.s.</i> $t(29) = -0.27304$, $p = .606620757191197$, $d = -0.049849817296457$
Whole Word: Written and Spoken	<i>n.s.</i> $t(29) = -0.42051$, $p = .661393903275375$, $d = -0.076774402915877$	<i>n.s.</i> $t(29) = -0.78274$, $p = .779936332342702$, $d = -0.142907435815172$
Repeated Exposure	<i>n.s.</i> $t(29) = 1.505199$, $p = .0715439550086867$, $d = 0.274810540797352$	<i>n.s.</i> $t(29) = 0$, $p = .50$, $d = 0$
Control: Untreated Unexposed	<i>n.s.</i> $t(29) = 0.140504$, $p = .444616665000135$, $d = 0.025652368718617$	<i>n.s.</i> $t(29) = -1.09691$, $p = .859147605696971$, $d = -0.200267954849917$

RT for Correct Response

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 0.945994$, $p = .17598379651844$, $d = 0.172714146315891$	<i>n.s.</i> $t(29) = -0.13315$, $p = .552502795018836$, $d = -0.0243097535364618$
Whole Word: Written Only	<i>*</i> $t(29) = 1.76585$, $p = .0439743123355941$, $d = 0.322398674874484$	<i>n.s.</i> $t(29) = 0.14078$, $p = .44450842368042$, $d = 0.0257028473269375$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 1.232163$, $p = .113892653534906$, $d = 0.224961195324357$	<i>n.s.</i> $t(29) = -0.21978$, $p = .586210580850434$, $d = -0.0401269590397219$
Whole Word: Written and Spoken	<i>n.s.</i> $t(29) = -0.11503$, $p = .545391337322239$, $d = -0.0210007660135963$	<i>n.s.</i> $t(29) = -0.52208$, $p = .697211274249045$, $d = -0.095318994567815$
Repeated Exposure	<i>*</i> $t(29) = 1.796286$, $p = .0414371735919744$, $d = 0.327955481022587$	<i>n.s.</i> $t(29) = -0.09195$, $p = .53631461991943$, $d = -0.0167875234106931$
Control: Untreated Unexposed	<i>n.s.</i> $t(29) = -0.1629$, $p = .564138308851435$, $d = -0.0297421890187472$	<i>n.s.</i> $t(29) = -1.19618$, $p = .879341997625377$, $d = -0.218391172302058$

† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

P5:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 0.116711$, $p = .453946864038794$, $d = 0.0213083852871282$	<i>n.s.</i> $t(29) = 0.384982$, $p = .351529895691731$, $d = 0.0702877997027422$
Whole Word: Written Only	<i>*</i> $t(29) = 2.19173$, $p = .0182889792500825$, $d = 0.400153344839829$	<i>n.s.</i> $t(29) = 1.394243$, $p = .0869177107921513$, $d = 0.2545527089295$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.644572$, $p = .262135549093508$, $d = 0.117682167917039$	<i>n.s.</i> $t(29) = -0.69407$, $p = .753418145650259$, $d = -0.12671852991977$
Whole Word: Written and Spoken	<i>n.s.</i> $t(29) = 1.458752$, $p = .0776902274590952$, $d = 0.26633039238097$	<i>n.s.</i> $t(29) = 0.350792$, $p = .364139895913724$, $d = 0.0640455990315034$
Repeated Exposure	<i>**</i> $t(29) = 2.719269$, $p = .00546687541972152$, $d = 0.496468270579197$	<i>n.s.</i> $t(29) = 1.05233$, $p = .1506690162914$, $d = 0.19212830504805$
Control: Untreated Unexposed	<i>**</i> $t(29) = 2.626415$, $p = .00681914889009054$, $d = 0.479515667190245$	<i>n.s.</i> $t(29) = 0.290349$, $p = .386807865032603$, $d = 0.053010164302517$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 0.341271$, $p = .367680007744064$, $d = 0.0623072664178145$	<i>n.s.</i> $t(29) = -0.11355$, $p = .544813268291096$, $d = -0.0207321166559166$
Whole Word: Written Only	<i>**</i> $t(29) = 2.587772$, $p = .00746873638629833$, $d = 0.47246039994537$	$\dagger$ $t(29) = 1.531795$, $p = .068205151017444$, $d = 0.279666287198749$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.563282$, $p = .288786292826624$, $d = 0.102840759117964$	<i>n.s.</i> $t(29) = -0.68063$, $p = .749248266586311$, $d = -0.124265681229438$
Whole Word: Written and Spoken	<i>*</i> $t(29) = 2.426359$, $p = .0108496560581442$, $d = 0.442990520783187$	<i>n.s.</i> $t(29) = 0.528059$, $p = .30073812556406$, $d = 0.0964099154385678$
Repeated Exposure	<i>**</i> $t(29) = 2.719269$, $p = .00546687541972152$, $d = 0.496468270579197$	<i>n.s.</i> $t(29) = 1.05233$, $p = .1506690162914$, $d = 0.19212830504805$
Control: Untreated Unexposed	$\dagger$ $t(29) = 1.628226$, $p = .0571471487301037$, $d = 0.297272008043423$	<i>n.s.</i> $t(29) = 0.25$, $p = .402174941540559$, $d = 0.0456435464587638$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	<i>n.s.</i> $t(29) = 0.175905$, $p = .430795717558231$, $d = 0.0321156486105259$	<i>n.s.</i> $t(29) = 0.07988$, $p = .468440539592973$, $d = 0.0145840989088653$
Whole Word: Written Only	<i>*</i> $t(29) = 2.392354$, $p = .0117208940284382$, $d = 0.436782040249252$	<i>n.s.</i> $t(29) = 1.389896$, $p = .0875692879310741$, $d = 0.253759072977842$
Initial Letter and Phoneme: Written and Spoken	<i>n.s.</i> $t(29) = 0.476175$, $p = .318758712029045$, $d = 0.0869373379125309$	<i>n.s.</i> $t(29) = -0.68829$, $p = .751631609130619$, $d = -0.125664809766473$
Whole Word: Written and Spoken	<i>*</i> $t(29) = 2.121163$, $p = .021289453445026$, $d = 0.387269540130221$	<i>n.s.</i> $t(29) = 0.551184$, $p = .292864834813333$, $d = 0.100631943058673$
Repeated Exposure	<i>*</i> $t(29) = 2.015548$, $p = .026598936983913$, $d = 0.36798696279793$	<i>n.s.</i> $t(29) = 0.580837$, $p = .282918627626671$, $d = 0.106045862982568$
Control: Untreated Unexposed	<i>n.s.</i> $t(29) = 1.302285$, $p = .101533688901688$, $d = 0.237763661167344$	<i>n.s.</i> $t(29) = 0.012482$, $p = .495063265100922$, $d = 0.00227889311441697$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

Appendix 4.2

Results for each participant and condition based on item-level WEST analysis on main naming outcomes (i.e., Picture naming accuracies (initial and final responses) and latencies)

Condition	Initial Letter: Orthography		Whole Word: Orthography		Initial Letter: Ortho and Phono		Whole Word: Ortho and Phono		Repeated Exposure		Control: Untreated	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P1												
Initial Naming Accuracy												
WEST	<i>n.s.</i>	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Final Naming Accuracy												
WEST	*	<i>n.s.</i>	***	<i>n.s.</i>	*	<i>n.s.</i>	***	*	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>
Follow-up (where nec.), e.g., 15 mins post- vs. BL							BL1 vs. BL2: <i>p</i> = .75 15 mins vs. BL2: <i>p</i> = .013* x1 week vs. BL2: <i>p</i> = .077					
RT for Correct Responses												
WEST	*	<i>n.s.</i>	**	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	**	†	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>
Follow-up (where nec.), e.g., 15 mins post- vs. BL							BL1 vs. BL2: <i>p</i> > .90 15 mins vs. BL2: <i>p</i> = .0249* x1 week vs. BL2: <i>p</i> = .20					

WEST: † $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

BL: Baseline

- Follow-up tests for comparisons against baseline 2 (after adjusting for Bonferroni's): † $.025 < p \leq .05$, * $.01 < p \leq .025$, ** $.001 < p \leq .01$, *** $p \leq .001$
- Comparing post-facilitation outcome against its repeated exposure counterpart (each time-point compared just once):
† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

–: Post hoc not required as effects of 'Repeated Exposure' were not significant, based on WEST

(continued)

Condition	Initial Letter: Orthography		Whole Word: Orthography		Initial Letter: Ortho and Phono		Whole Word: Ortho and Phono		Repeated Exposure		Control: Untreated	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P2												
Initial Naming Accuracy												
WEST	**	*	**	n.s.	n.s.	n.s.	*	n.s.	n.s.	n.s.	*	n.s.
Follow-Up (where nec.), e.g., 15 mins post- vs. BL	BL1 vs. BL2: $p = n.s.$ 15 mins vs. BL2: $p = .063$ x1 wk vs. BL2: $p = .063$											
Post-Facilitation vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p = .50$ 15 mins post- vs. BL2: $p = .69$ x1 week post-: $p > .90$											
Final Naming Accuracy												
WEST	**	n.s.	***	**	**	*	**	n.s.	**	†	*	n.s.
Follow-up (where nec.), e.g., 15 mins post- vs. BL			BL1 vs. BL2: $p = .73$ 15 mins vs. BL2: $p = .002^{**}$ x1 wk vs. BL2: $p = .004^{**}$		BL1 vs. BL2: $p = .73$ 15 mins vs. BL2: $p = .016^{*}$ x1 wk vs. BL2: $p = .008^{**}$				BL1 vs. BL2: $p = .73$ 15 mins vs. BL2: $p = .07$ x1 wk vs. BL2: $p = .031^{\dagger}$			
Post-Facilitation vs. Rep. Exp. Counterpart (where nec.)			Final BLs: $p > .90$ 15 minutes post-: $p = .22$ x1 week post-: $p = .38$		Final BLs: $p > .90$ 15 minutes post-: $p > .90$ x1 week post-: $p = .63$							
RT for Correct Responses												
WEST	**	n.s.	***	**	**	*	***	n.s.	†	†	*	n.s.
Follow-up (where nec.), e.g., 15 mins post- vs. BL			BL1 vs. BL2: $p = .48$ 15 mins vs. BL2: $p = .002^{**}$ x1 wk vs. BL2: $p = .002^{**}$		BL1 vs. BL2: $p = .33$ 15 mins vs. BL2: $p = .009^{**}$ x1 wk vs. BL2: $p = .004^{**}$				BL1 vs. BL2: $p = .26$ 15 mins vs. BL2: $p = .028^{\dagger}$ 1 wk vs. BL2: $p > .05$			
Post-Facilitation vs. Rep. Exp. Counterpart (where nec.)			Final BLs: $p = .89$ 15 minutes post-: $p = .14$ x1 week post-: $p = .046^{*}$		Final BLs: $p = .69$ 15 minutes post-: $p = .64$ x1 week post-: $p = .24$							

WEST: † .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001Follow-up tests (after adjusting for Bonferroni's): † .025 < p ≤ .05, * .01 < p ≤ .025, ** .001 < p ≤ .01, *** p ≤ .001

BL: Baseline

Comparing post-facilitation outcome against its repeated exposure counterpart (each time point compared just once): † .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

(continued)

Condition	Initial Letter: Orthography		Whole Word: Orthography		Initial Letter: Ortho and Phono		Whole Word: Ortho and Phono		Repeated Exposure		Control: Untreated	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P3												
Initial Naming Accuracy												
WEST	**	**	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Follow-Up (where nec.), e.g., 15 mins post- vs. BL	BL1 vs. BL2: <i>p</i> = .63 15 mins vs. BL2: <i>p</i> = .008** x1 wk vs. BL2: <i>p</i> = .63											
Final Naming Accuracy												
WEST	**	***	***	*	<i>n.s.</i>	*	*	†	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Follow-up (where nec.), e.g., 15 mins post- vs. BL	BL1 vs. BL2: <i>p</i> > .90 15 mins vs. BL2: <i>p</i> = .00098*** x1 wk vs. BL2: <i>p</i> = .22		BL1 vs. BL2: <i>p</i> > .90 15 mins vs. BL2: <i>p</i> = .039 [†] x1 wk vs. BL2: <i>p</i> = .23				BL1 vs. BL2: <i>p</i> > .90 15 mins vs. BL2: <i>p</i> = .109 x1 wk vs. BL2: <i>p</i> > .90					
RT for Correct Responses												
WEST	**	**	**	†	<i>n.s.</i>	*	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Follow-up (where nec.), e.g., 15 mins post- vs. BL	BL1 vs. BL2: <i>p</i> = .22 15 mins vs. BL2: <i>p</i> = .004** x1 wk vs. BL2: <i>p</i> = .26		BL1 vs. BL2: <i>p</i> = .89 15 mins vs. BL2: <i>p</i> = .055 x1 wk vs. BL2: <i>p</i> = .11									

WEST: † .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

BL: Baseline

- Follow-up tests for comparisons against baseline 2 (after adjusting for Bonferroni's): † .025 < $p \leq .05$, * .01 < $p \leq .025$, ** .001 < $p \leq .01$, *** $p \leq .001$
- Comparing post-facilitation outcome against its repeated exposure counterpart (each time-point compared just once):
† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

–: Post hoc not required as effects of 'Repeated Exposure' were not significant, based on WEST

(continued)

Condition	Initial Letter: Orthography		Whole Word: Orthography		Initial Letter: Ortho and Phono		Whole Word: Ortho and Phono		Repeated Exposure		Control: Untreated	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P4												
Initial Naming Accuracy												
WEST	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Final Naming Accuracy												
WEST	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
RT for Correct Responses												
WEST	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
P5												
Initial Naming Accuracy												
WEST	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	**	<i>n.s.</i>	**	<i>n.s.</i>
Final Naming Accuracy												
WEST	<i>n.s.</i>	<i>n.s.</i>	**	†	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	**	<i>n.s.</i>	†	<i>n.s.</i>
Follow-up (where nec.): e.g., 15 mins post- vs. BL			BL1 vs. BL2: <i>p</i> > .90 15 mins vs. BL2: <i>p</i> = .065 x1 wk vs. BL2: <i>p</i> = .0703									
RT for Correct Responses												
WEST	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	*	<i>n.s.</i>	*	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>

WEST: † .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

BL: Baseline

- Follow-up tests for comparisons against baseline 2 (after adjusting for Bonferroni's): † .025 < p ≤ .05, * .01 < p ≤ .025, ** .001 < p ≤ .01, *** p ≤ .001
- Comparing post-facilitation outcome against its repeated exposure counterpart (each time-point compared just once):
† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

–: Post hoc not required as effects of 'Repeated Exposure' were not significant, based on WEST

Appendix 5.1

Brief description of baseline tasks administered during the orthographic cueing therapy study

Language assessments:

Name of task/test	Description
Oromotor speech screeners	
Apraxia Battery for Adults (ABA-2, Dhabul, 2000): Subtests 1 to 3	Subtest 1: Diadochokinetic Rate - Rapid repetition of different syllables of varying length (e.g., /pʌ, tʌkʌ/) for 1, 3 and 5 seconds Subtest 2b: Increasing word length - Repeating the spoken target word produced by tester - Words increase in syllable length and complexity (e.g., 'beauty', 'beautiful', 'beautifully') Subtest 3: Limb apraxia and oral apraxia - Assess oro-masculature integrity and overall limb/oromotor coordination
Spoken language processing	
PALPA 23. Spoken letter sound-written letter matching	Choose the correct printed alphabet (choice of four) that corresponds to the spoken phoneme given by the tester: e.g., /b/ for 'b'
PALPA 14. Rhyme judgement: Picture selection ¹	Indicate 'Yes' if the target words corresponding to the two pictures rhyme: e.g., pictures of 'cat' and 'hat'
CAT 12. Repetition of words	Verbally repeat the spoken words (single morphemes, e.g., 'table') given by the tester
CAT 13. Repetition of complex words	Verbally repeat the spoken words (multimorphemic, e.g., 'unthinkable') given by the tester
PALPA 8. Nonword repetition task	Verbally repeat the spoken pseudo-word given by the tester, e.g., 'vater'
Written language processing	
PALPA 18. Letter discrimination: Mirror reversal	Indicate 'Yes' for letters printed in the correct orientation
PALPA 22. Letter sounding	Produce the sound corresponding to the printed alphabet, e.g., /p/ for 'p'
PALPA 22. Letter naming	Say out the name of the printed alphabet: e.g., /pi/ for 'p'
PALPA 25. Visual lexical decision task	Indicate 'Yes' for real words printed on the sheet, and 'No' to the pseudo-words

(continued)

Name of task/test	Description
Written language processing PALPA 35a. Spelling-sound regularity reading task: Regular and exception words PALPA 36. Reading nonwords	• Read aloud the words printed on the sheet, e.g., 'rub', 'quay' • Read aloud the pseudo-words printed on the sheet (e.g., 'grest')
Semantic processing Pyramid and Palm Trees: Three pictures version PALPA 47: Spoken word-picture matching PALPA 48: Written word-picture matching PALPA 50. Written synonym judgement	• Choose the picture (choice of two) that is most related to the target picture in the middle, e.g., 'banana' with target picture, 'monkey' • Choose the correct picture (choice of five) that corresponds to the spoken target word given by the tester • Choose the correct picture (choice of five) that corresponds to the written target word printed in the middle • Indicate 'Yes' if the two written words are close in meaning. Indicate 'No' if otherwise.
Naming Boston Naming Test (BNT) CAT 25. Writing picture names	• Name the target item represented in the picture (e.g., 'bed') • Write out the target word for the picture (all stimuli are nouns; e.g., 'hat')

¹ 'PALPA 14. Rhyme judgement: Picture selection' neither involves spoken stimuli nor spoken output. However, it involves the silent comparison and thus, activation of how the target words sound, and thus placed under 'spoken language processing'.

Cognitive assessments:

Name of task/test	Description
Attention	
Test of Everyday Attention (TEA) Subtest 2: Elevator counting	• Mono-pitched tone sounds, resembling what may be heard in an elevator, are presented at irregular tempos. Listen and count the number of tone sounds.
TEA Subtest 3: Elevator counting with distraction	• High and low-pitched tone sounds, resembling what may be heard in an elevator, are presented at irregular tempos. Listen and count the number of low-pitched tones sounds. Ignore the high-pitched tones.
Memory	
CAT 4: Recognition memory	• Choose the correct picture (choice of four) that corresponds to an earlier picture previously presented about 5 minutes ago
Corsi Block Tapping Test	• Watch the computerized sequence in which the squares are indicated on the screen • Tap the squares in the same sequence that was just shown • Squares arranged according to Claessen, van der Ham, and van Zandvoort (2015), and sequence followed Smirni, Villardita and Zappalá (1983). Squares indicated at one block per second, following Claessen and colleagues. • Online assessment of this task required support from the caregiver, who read aloud the alphabet of the square that the participant simultaneously pointed at.
Wechsler Memory Scale (WMS) 5a: Digits forward recall	• Verbally repeat the numbers in the same order spoken by the tester • Number span ranges from 4 to 8
WMS 5b: Digits backward recall	• Verbally repeat the numbers in the reverse order spoken by the tester • Number span ranges from 3 to 7

(continued)

Name of task/test	Description
Executive function and nonverbal reasoning	
Brixton Spatial Anticipation Test	• Assesses 'concept/rule attainment' • For the stimulus, a filled circle will change its position along a 2x5 array, according to fixed rules • Point to the next position within the array that the filled circle will move to next
Raven's Coloured Matrices (Sets A, A _B , B)	• Choose the correct picture (choice of 6) that completes the target pattern shown at the top

Appendix 5.2a

Post-therapy Questionnaire: For the Participant with Aphasia (Note that rating responses were reverse coded for analyses)

(1) Is it easy to express what you want?

Very Easy

Very Difficult



1



2



3



4



5



6

(2) Is it difficult to say out the correct word?

Not Difficult at all

Very Difficult



1



2



3



4



5



6

Questions (3) [ease of skype] and (4) [rate online therapy experience] are only asked after the respective therapies on Weeks 10 and 13, and Question (5) [preferred therapy], after both therapies on Week 13:

(3) How easy or difficult was using skype in therapy?

Very Easy

Very Difficult



1



2



3



4



5



6

(4) In your last two weeks, how was your experience with our therapy on the internet?

Very Good

Neutral

Very Poor



1



2



3



4

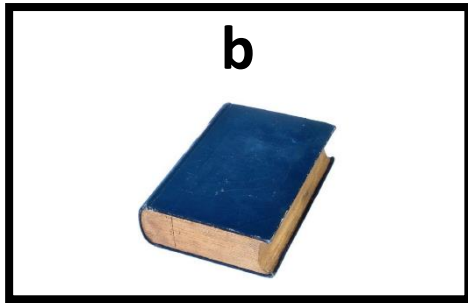


5

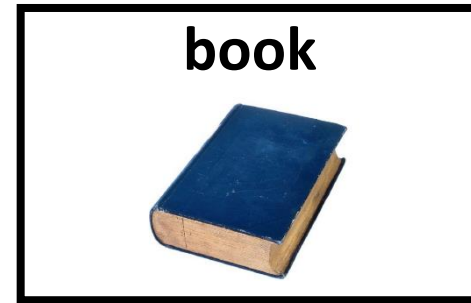


6

(5a) You went through two therapies. Which therapy do you prefer?



(a) First Letter



(b) Whole Word

(5b) Why?

Appendix 5.2b

Post-therapy Questionnaire: For the Caregiver (Note that rating responses were reverse coded for analyses)

(3) Is it easy for [Participant's Name] to express what (s)he wants?

Very Easy

Very Difficult



1



2



3



4



5



6

(4) Is it difficult for [Participant's Name] to say out the correct word?

Not Difficult at all

Very Difficult



1

2

3

4

5

6

Question (3) [ease of skype] is only asked after the respective therapies on Weeks 10 and 13:

(3) How easy or difficult was running skype for therapy?

Very Easy

Very Difficult



1

2

3

4

5

6

Appendix 5.3

Changes in performance on all three main outcome measures using mean proportion (*SD*) across participants and treated items, as well as changes in performance on accuracy measures using average raw scores (*SD*) across participants.

Time Post-Therapy:	<u>Initial Naming Accuracy</u>				<u>Final Naming Accuracy</u>			
	<u>1 Week</u>		<u>4 Weeks</u>		<u>1 Week</u>		<u>4 Weeks</u>	
Change in Performance from the Baseline Immediately Prior to Therapy:	Δ Mean Proportion	Δ Mean Raw Score	Δ Mean Proportion	Δ Mean Raw Score	Δ Mean Proportion	Δ Mean Raw Score	Δ Mean Proportion	Δ Mean Raw Score
<u>Condition:</u>								
(1) Initial Letter Orthographic Cue	0.329* (<i>SD</i> = 0.186)	7.40* (<i>SD</i> = 5.03)	0.330* (<i>SD</i> = 0.131)	7.20* (<i>SD</i> = 3.27)	0.425* (<i>SD</i> = 0.229)	9.60* (<i>SD</i> = 6.58)	0.374* (<i>SD</i> = 0.183)	8.20* (<i>SD</i> = 4.76)
(2) Repeated Exposure: Initial Letter Ortho. Cue	0.221 [†] (<i>SD</i> = 0.199)	4.60 [†] (<i>SD</i> = 3.85)	0.232* (<i>SD</i> = 0.122)	5.00* (<i>SD</i> = 2.74)	0.284* (<i>SD</i> = 0.145)	5.80* (<i>SD</i> = 2.95)	0.267* (<i>SD</i> = 0.0956)	5.60* (<i>SD</i> = 2.30)
(3) Whole Word Orthographic Cue	0.463 [†] (<i>SD</i> = 0.290)	10.20 [†] (<i>SD</i> = 6.76)	0.327* (<i>SD</i> = 0.122)	7.20* (<i>SD</i> = 2.59)	0.612* (<i>SD</i> = 0.211)	13.00* (<i>SD</i> = 5.43)	0.478* (<i>SD</i> = 0.0797)	10.00* (<i>SD</i> = 1.87)
(4) Repeated Exposure: Whole Word Ortho. Cue	0.192* (<i>SD</i> = 0.139)	4.20* (<i>SD</i> = 2.86)	0.195 [†] (<i>SD</i> = 0.130)	4.00 [†] (<i>SD</i> = 2.55)	0.224 [†] (<i>SD</i> = 0.168)	4.80 [†] (<i>SD</i> = 3.70)	0.251* (<i>SD</i> = 0.0859)	5.00* (<i>SD</i> = 1.41)
(5a) Unseen Control: Initial Letter Cue as Therapy 1 (<i>n</i> = 3)	0.138 (<i>SD</i> = 0.0893)	3.33 (<i>SD</i> = 2.31)	0.110 (<i>SD</i> = 0.0100)	2.67 (<i>SD</i> = 0.58)	0.176 (<i>SD</i> = 0.0932)	4.33 (<i>SD</i> = 2.52)	0.161 (<i>SD</i> = 0.0611)	4.00 (<i>SD</i> = 2.00)
(5b) Unseen Control: Whole Word Cue as Therapy 1 (<i>n</i> = 2)	0.0714 (<i>SD</i> = 0.101)	1.50 (<i>SD</i> = 2.12)	0.179 (<i>SD</i> = 0.0168)	3.50 (<i>SD</i> = 0.71)	0.158 (<i>SD</i> = 0.0118)	3.00 (<i>SD</i> = 0)	0.208 (<i>SD</i> = 0.0589)	4.00 (<i>SD</i> = 1.41)

Cells highlighted in yellow refer to conditions whose change in performance at post-therapy, compared to the baseline performance just prior to therapy, was significant based on post hoc testing.

Cells with bold and red letterings refer to conditions whose results remained significant when compared against the corresponding repeated exposure conditions, based on post hoc testing.

[†] .05 < *p* < .07, * .01 < *p* < .05, ** .001 < *p* < .01, *** *p* < .001

(continued)

<u>RT (ms)</u>				
Time Post-Therapy:	<u>1 Week</u>		<u>4 Weeks</u>	
Change in Performance from the Baseline Immediately Prior to Therapy:	Δ Mean Proportion	Δ Mean Raw Value	Δ Mean Proportion	Δ Mean Raw Value
<u>Condition:</u>				
(1) Initial Letter Orthographic Cue	0.259* (SD = 0.154)	2507.97* (SD = 1667.15)	0.239* (SD = 0.140)	2287.88* (SD = 1446.88)
(2) Repeated Exposure: Initial Letter Ortho. Cue	0.222* (SD = 0.133)	2092.62* (SD = 1190.58)	0.176* (SD = 0.0823)	1686.78* (SD = 853.04)
(3) Whole Word Orthographic Cue	0.378* (SD = 0.199)	3647.54* (SD = 2008.11)	0.292* (SD = 0.0804)	2803.69* (SD = 779.15)
(4) Repeated Exposure: Whole Word Ortho. Cue	0.182* (SD = 0.125)	1752.66* (SD = 1664.51)	0.179* (SD = 0.103)	1673.90* (SD = 864.75)
(5a) Unseen Control: Initial Letter Cue as Therapy 1 ($n = 3$)	0.112 (SD = 0.0783)	1152.42 (SD = 822.28)	0.0941 (SD = 0.0616)	974.51 (SD = 657.34)
(5b) Unseen Control: Whole Word Cue as Therapy 1 ($n = 2$)	0.0827 (SD = 0.0566)	790.25 (SD = 619.24)	0.159 (SD = 0.0406)	1473.97 (SD = 567.61)

Cells highlighted in yellow refer to conditions whose change in performance at post-therapy, compared to the baseline performance just prior to therapy, was significant based on post hoc testing.

Cells with bold and red letterings refer to conditions whose results remained significant when compared against the corresponding repeated exposure conditions, based on post hoc testing.

[†] .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

Appendix 6.1
Individual results based on WEST-Trend and ROC for
each main naming outcome measure

P1:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(25) = 3.923612728$, $p = .00030134$, $d = 0.769483764$	<i>n.s.</i> $t(25) = 1.282568341$, $p = .105704122$, $d = 0.251532346$
Whole Word (Written Only)	*** $t(25) = 4.541430052$, $p = .0000611607$, $d = 0.89064771$	* $t(25) = 1.791093316$, $p = .042697324$, $d = 0.351262299$
Repeated Exposure: Initial Letter (Written Only)	*** $t(25) = 4.795370526$, $p = .0000316602$, $d = 0.940449534$	* $t(25) = 2.481510576$, $p = .01008085$, $d = 0.486664263$
Repeated Exposure: Whole Word (Written Only)	*** $t(25) = 4.919011134$, $p = .0000229819$, $d = 0.964697452$	* $t(25) = 1.901007852$, $p = .034444437$, $d = 0.372818313$
Unseen Control: Untreated, Unexposed	*** $t(25) = 3.521476061$, $p = .00083643$, $d = 0.690618275$	<i>n.s.</i> $t(25) = -0.420018145$, $p = .660969921$, $d = -0.082372335$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(25) = 4.128167638$, $p = .000178132$, $d = 0.809600282$	<i>n.s.</i> $t(25) = -0.028666478$, $p = .511320895$, $d = -0.005621959$
Whole Word (Written Only)	*** $t(25) = 5.127255972$, $p = .0000134117$, $d = 1.005537625$	* $t(25) = 1.919869896$, $p = .033177322$, $d = 0.376517464$
Repeated Exposure: Initial Letter (Written Only)	*** $t(25) = 4.936190706$, $p = .000021982$, $d = 0.968066644$	† $t(25) = 1.535858123$, $p = .068566853$, $d = 0.301206559$
Repeated Exposure: Whole Word (Written Only)	*** $t(25) = 5.060249395$, $p = .0000159467$, $d = 0.992396554$	<i>n.s.</i> $t(25) = 0.762079675$, $p = .226571369$, $d = 0.149456121$
Unseen Control: Untreated, Unexposed	** $t(25) = 3.376138617$, $p = .001202748$, $d = 0.662115257$	<i>n.s.</i> $t(25) = 0.67997673$, $p = .251383907$, $d = 0.133354408$

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	*** $t(25) = 4.840256962$, $p = .0000281832$, $d = 0.949252488$	<i>n.s.</i> $t(25) = 0.116316707$, $p = .454165445$, $d = 0.022811583$
Whole Word: Written Only	*** $t(25) = 4.97291177$, $p = .0000199887$, $d = 0.975268237$	* $t(25) = 2.437702192$, $p = .011115362$, $d = 0.478072733$
Repeated Exposure: Initial Letter (Written Only)	*** $t(25) = 5.234943654$, $p = .0000101584$, $d = 1.026656917$	* $t(25) = 2.185254433$, $p = .019226452$, $d = 0.428563654$
Repeated Exposure: Whole Word (Written Only)	*** $t(25) = 4.677630412$, $p = .0000429628$, $d = 0.917358798$	* $t(25) = 1.843672375$, $p = .038558212$, $d = 0.361573901$
Unseen Control: Untreated, Unexposed	*** $t(25) = 3.641387598$, $p = .00061826$, $d = 0.714134862$	<i>n.s.</i> $t(25) = 0.108255213$, $p = .457328713$, $d = 0.021230594$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

P2:**Accuracy of Initial Responses**

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	** $t(23) = 3.428292076$, $p = .001147741$, $d = 0.69979719$	** $t(23) = 2.969838544$, $p = .003429251$, $d = 0.606215754$
Whole Word (Written Only)	*** $t(23) = 5.361902647$, $p = .00000954695$, $d = 1.094493795$	*** $t(23) = 4.725255398$, $p = .000046162$, $d = 0.964538719$
Repeated Exposure: Initial Letter (Written Only)	* $t(23) = 1.983892355$, $p = .029663759$, $d = 0.404960331$	† $t(23) = 1.704935396$, $p = 0.050839543$, $d = 0.34801848$
Repeated Exposure: Whole Word (Written Only)	** $t(23) = 2.942280678$, $p = .003657408$, $d = 0.600590529$	<i>n.s.</i> $t(23) = 1.221162017$, $p = .11719768$, $d = 0.249268653$
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $t(23) = 1.251980351$, $p = .111579917$, $d = 0.255559419$	<i>n.s.</i> $t(23) = 0.983274484$, $p = .167851718$, $d = 0.200710064$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(23) = 4.258096624$, $p = .000147966$, $d = 0.869180334$	*** $t(23) = 3.574038567$, $p = .000804572$, $d = 0.729547567$
Whole Word (Written Only)	*** $t(23) = 7.467845285$, $p = .0000000684224$, $d = 1.524367536$	*** $t(23) = 5.079632939$, $p = .000019147$, $d = 1.036875732$
Repeated Exposure: Initial Letter (Written Only)	* $t(23) = 2.243390522$, $p = .017399088$, $d = 0.457930173$	* $t(23) = 2.314204129$, $p = .01496882$, $d = 0.47238494$
Repeated Exposure: Whole Word (Written Only)	** $t(23) = 3.139608711$, $p = .002297331$, $d = 0.640869944$	** $t(23) = 2.967478714$, $p = .003448243$, $d = 0.605734056$
Unseen Control: Untreated, Unexposed	† $t(23) = 1.544889762$, $p = .068011099$, $d = 0.315349302$	<i>n.s.</i> $t(23) = 0.859224187$, $p = .199541929$, $d = 0.175388403$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	*** $t(23) = 4.135618594$, $p = .000200758$, $d = 0.844179611$	** $t(23) = 3.400337588$, $p = .001228256$, $d = 0.694091004$
Whole Word: Written Only	*** $t(23) = 6.015418874$, $p = .00000195488$, $d = 1.227892236$	*** $t(23) = 5.432276141$, $p = .00000803345$, $d = 1.108858724$
Repeated Exposure: Initial Letter (Written Only)	* $t(23) = 2.472934398$, $p = .010610708$, $d = 0.50478562$	** $t(23) = 2.713288949$, $p = .00619952$, $d = 0.553847788$
Repeated Exposure: Whole Word (Written Only)	** $t(23) = 3.177110168$, $p = .002101065$, $d = 0.648524897$	* $t(23) = 2.227545003$, $p = .017989899$, $d = 0.45469572$
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $t(23) = 1.420540692$, $p = .084432884$, $d = 0.289966655$	<i>n.s.</i> $t(23) = 0.987823197$, $p = .166758987$, $d = 0.201638566$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

P3:**Accuracy of Initial Responses**

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	† $t(24) = 1.710204247$, $p = .050063535$, $d = 0.342040849$	<i>n.s.</i> $t(24) = 1.000538486$, $p = .163515845$, $d = 0.200107697$
Whole Word (Written Only)	*** $t(24) = 3.636259993$, $p = .00065675$, $d = 0.727251999$	*** $t(24) = 4.420732424$, $p = .0000905999$, $d = 0.884146485$
Repeated Exposure: Initial Letter (Written Only)	** $t(24) = 3.25875268$, $p = .001665163$, $d = 0.651750536$	<i>n.s.</i> $t(24) = 0.479808115$, $p = .317853836$, $d = 0.095961623$
Repeated Exposure: Whole Word (Written Only)	** $t(24) = 2.889141748$, $p = .004031616$, $d = 0.57782835$	<i>n.s.</i> $t(24) = 0.271825434$, $p = .394040469$, $d = 0.054365087$
Unseen Control: Untreated, Unexposed	** $t(24) = 2.941262223$, $p = .00356604$, $d = 0.588252445$	<i>n.s.</i> $t(24) = 0.103954564$, $p = .459034623$, $d = 0.020790913$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	* $t(24) = 2.247600584$, $p = .017031048$, $d = 0.449520117$	<i>n.s.</i> $t(24) = 0.684853628$, $p = .24999876$, $d = 0.136970726$
Whole Word (Written Only)	*** $t(24) = 3.917334539$, $p = .000324574$, $d = 0.783466908$	*** $t(24) = 3.764637006$, $p = .000476468$, $d = 0.752927401$
Repeated Exposure: Initial Letter (Written Only)	*** $t(24) = 3.764892383$, $p = .000476163$, $d = 0.752978477$	<i>n.s.</i> $t(24) = 0.929000841$, $p = .181069394$, $d = 0.185800168$
Repeated Exposure: Whole Word (Written Only)	** $t(24) = 3.074171006$, $p = .002599811$, $d = 0.614834201$	<i>n.s.</i> $t(24) = -0.045980958$, $p = .518147053$, $d = -0.009196192$
Unseen Control: Untreated, Unexposed	* $t(24) = 2.267786838$, $p = .016313301$, $d = 0.453557368$	<i>n.s.</i> $t(24) = 0.11483385$, $p = .454766082$, $d = 0.02296677$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	* $t(24) = 2.106235494$, $p = .02291358$, $d = 0.421247099$	<i>n.s.</i> $t(24) = 0.532277772$, $p = .299713642$, $d = 0.106455554$
Whole Word: Written Only	** $t(24) = 2.93361557$, $p = .003630983$, $d = 0.586723114$	** $t(24) = 3.219247408$, $p = .001832846$, $d = 0.643849482$
Repeated Exposure: Initial Letter (Written Only)	** $t(24) = 3.396397716$, $p = .001189279$, $d = 0.679279543$	<i>n.s.</i> $t(24) = 0.84668716$, $p = .202766682$, $d = 0.169337432$
Repeated Exposure: Whole Word (Written Only)	** $t(24) = 2.991453729$, $p = .00316649$, $d = 0.598290746$	<i>n.s.</i> $t(24) = -0.004152806$, $p = 0.501639562$, $d = -0.000830561$
Unseen Control: Untreated, Unexposed	* $t(24) = 1.740108331$, $p = .04732489$, $d = 0.348021666$	<i>n.s.</i> $t(24) = 0.068337258$, $p = .473041666$, $d = 0.013667452$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

P4:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	** $t(27) = 3.055844718$, $p = .00250417$, $d = 0.577500369$	* $t(27) = 1.906403301$, $p = .033646762$, $d = 0.360276359$
Whole Word (Written Only)	* $t(27) = 2.282218388$, $p = .015288355$, $d = 0.431298735$	<i>n.s.</i> $t(27) = 0.275315458$, $p = .392584958$, $d = 0.052029731$
Repeated Exposure: Initial Letter (Written Only)	* $t(27) = 2.152399863$, $p = .020232383$, $d = 0.40676534$	<i>n.s.</i> $t(27) = -0.719496529$, $p = .760991454$, $d = -0.135972063$
Repeated Exposure: Whole Word (Written Only)	<i>n.s.</i> $t(27) = 1.013054166$, $p = .160013561$, $d = 0.191449242$	<i>n.s.</i> $t(27) = 0.799579352$, $p = .21546661$, $d = 0.151106294$
Unseen Control: Untreated, Unexposed	<i>n.s.</i> $t(27) = 1.384103607$, $p = .088830889$, $d = 0.261570995$	† $t(27) = 1.566176017$, $p = .064476972$, $d = 0.295979446$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(27) = 5.749278012$, $p = .00000205315$, $d = 1.086511417$	*** $t(27) = 3.850393701$, $p = .000328295$, $d = 0.727656013$
Whole Word (Written Only)	*** $t(27) = 5.988831263$, $p = .00000108943$, $d = 1.131782726$	** $t(27) = 2.892591627$, $p = .003731853$, $d = 0.546648435$
Repeated Exposure: Initial Letter (Written Only)	** $t(27) = 3.275348601$, $p = .001447765$, $d = 0.618982704$	<i>n.s.</i> $t(27) = 0.711601429$, $p = .241407528$, $d = 0.13448003$
Repeated Exposure: Whole Word (Written Only)	* $t(27) = 2.28919851$, $p = .015056483$, $d = 0.432617854$	* $t(27) = 1.815356537$, $p = .040296229$, $d = 0.343070138$
Unseen Control: Untreated, Unexposed	** $t(27) = 3.134240912$, $p = .002062022$, $d = 0.592315857$	* $t(27) = 1.716616206$, $p = .048751228$, $d = 0.32440997$

† .05 < $p \leq .07$, * .01 < $p \leq .05$, ** .001 < $p \leq .01$, *** $p \leq .001$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	*** $t(27) = 4.605246321$, $p = .0000439903$, $d = 0.870309749$	*** $t(27) = 3.579906934$, $p = .000664657$, $d = 0.676538819$
Whole Word: Written Only	*** $t(27) = 4.936880247$, $p = .0000180431$, $d = 0.932982671$	* $t(27) = 2.012895443$, $p = .027097934$, $d = 0.380401483$
Repeated Exposure: Initial Letter (Written Only)	** $t(27) = 3.023106556$, $p = .002714435$, $d = 0.571313438$	<i>n.s.</i> $t(27) = 0.202490423$, $p = .420525728$, $d = 0.038267093$
Repeated Exposure: Whole Word (Written Only)	* $t(27) = 1.742396884$, $p = .046410307$, $d = 0.32928206$	** $t(27) = 1.833757763$, $p = 0.038867904$, $d = 0.346547643$
Unseen Control: Untreated, Unexposed	** $t(27) = 2.621293619$, $p = .007106511$, $d = 0.495377931$	† $t(27) = 1.651475082$, $p = .055113331$, $d = 0.312099455$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

P5:

Accuracy of Initial Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(27) = 5.472413213$, $p = .00000429186$, $d = 1.034188888$	*** $t(27) = 5.05366068$, $p = .0000131833$, $d = 0.955052098$
Whole Word (Written Only)	*** $t(27) = 6.74056154$, $p = .000000154194$, $d = 1.273846395$	*** $t(27) = 5.726746705$, $p = .00000217973$, $d = 1.0822534$
Repeated Exposure: Initial Letter (Written Only)	*** $t(27) = 5.2516907$, $p = .00000774818$, $d = 0.992476254$	* $t(27) = 1.718271346$, $p = .048597999$, $d = 0.324722762$
Repeated Exposure: Whole Word (Written Only)	** $t(27) = 3.262642047$, $p = .001494912$, $d = 0.616581391$	* $t(27) = 1.712541048$, $p = 0.049130235$, $d = 0.323639837$
Unseen Control: Untreated, Unexposed	* $t(27) = 1.903760753$, $p = .033825444$, $d = 0.359776965$	* $t(27) = 2.404844784$, $p = .011652358$, $d = 0.454472946$

Accuracy of Final Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter (Written Only)	*** $t(27) = 6.223834219$, $p = .000000587755$, $d = 1.17619411$	*** $t(27) = 5.229625854$, $p = .00000822042$, $d = 0.98830639$
Whole Word (Written Only)	*** $t(27) = 8.744680851$, $p = .00000000116191$, $d = 1.652589345$	*** $t(27) = 7.063173948$, $p = .0000000678649$, $d = 1.33481441$
Repeated Exposure: Initial Letter (Written Only)	*** $t(27) = 4.876569881$, $p = .0000212186$, $d = 0.921585083$	*** $t(27) = 3.633932485$, $p = .000577848$, $d = 0.686748688$
Repeated Exposure: Whole Word (Written Only)	** $t(27) = 3.317965653$, $p = .001299873$, $d = 0.62703657$	** $t(27) = 2.950212705$, $p = .003244617$, $d = 0.557537795$
Unseen Control: Untreated, Unexposed	* $t(27) = 1.759899504$, $p = .044875931$, $d = 0.332589744$	** $t(27) = 2.678043109$, $p = .006223878$, $d = 0.506102576$

† $.05 < p \leq .07$, * $.01 < p \leq .05$, ** $.001 < p \leq .01$, *** $p \leq .001$

RT for Correct Responses

Cue Condition	WEST-Trend	WEST-ROC
Initial Letter: Written Only	*** $t(27) = 6.938134429$, $p = .0000000931472$, $d = 1.311184162$	*** $t(27) = 5.045152165$, $p = .0000134881$, $d = 0.95344414$
Whole Word: Written Only	*** $t(27) = 7.660371091$, $p = .0000000153531$, $d = 1.447674061$	*** $t(27) = 7.000607438$, $p = .0000000794985$, $d = 1.322990451$
Repeated Exposure: Initial Letter (Written Only)	*** $t(27) = 4.9202707$, $p = .0000188669$, $d = 0.929843761$	*** $t(27) = 3.790213802$, $p = .000384441$, $d = 0.716283081$
Repeated Exposure: Whole Word (Written Only)	*** $t(27) = 3.423189785$, $p = .000994505$, $d = 0.646922062$	** $t(27) = 2.822995426$, $p = 0.004412767$, $d = 0.533495989$
Unseen Control: Untreated, Unexposed	* $t(27) = 1.911335481$, $p = .033315456$, $d = 0.361208454$	** $t(27) = 3.030271916$, $p = .002667024$, $d = 0.572667564$

† .05 < p ≤ .07, * .01 < p ≤ .05, ** .001 < p ≤ .01, *** p ≤ .001

Appendix 6.2

Results for each participant and condition based on item-level WEST analysis on main naming outcomes (i.e., Picture naming accuracies (initial and final responses) and latencies)

Condition		Initial Letter (Orthography)		Repeated Exposure: Initial Letter (Orthography)		Whole Word (Orthography)		Repeated Exposure: Whole Word (Orthography)		Unseen Control: Untreated, Unexposed	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	
P1											
Initial Naming Accuracy											
WEST Follow-up (where nec.)	***	<i>n.s.</i>	***	*	***	*	***	*	***	<i>n.s.</i>	
			Across x4 BLs: $\chi^2(3, N = 26) = 2.700, p = .440$, Kendall's $W = .035$		Across x3 BLs: $\chi^2(2, N = 26) = 3.500, p = .174$, Kendall's $W = .067$		Across x3 BLs: $\chi^2(2, N = 26) = 0.800, p = .670$, Kendall's $W = .015$				
			x1 week post-treatment vs. Final BL: $p = .002^{**}$		x1 week post-treatment vs. Final BL: $p = .065$		x1 week post-treatment vs. Final BL: $p = .021^{\dagger}$				
			x4 weeks post-treatment vs. Final BL: $p = .180$		x4 weeks post-treatment vs. Final BL: $p = .021^{\dagger}$		x4 weeks post-treatment vs. Final BL: $p = .039^{\dagger}$				
					x7 weeks post-treatment vs. Final BL: $p = .219$		x7 weeks post-treatment vs. Final BL: $p = .003^*$				
Treatment vs. Rep. Exp. Counterpart (where nec.)					Final BLs: $p > .90$						
					x1 week post-treatment: $p > .90$ x4 weeks post-treatment: $p = .774$ x7 weeks post-treatment: $p = .180$						
Final Naming Accuracy											
WEST Follow-up (where nec.)	***	<i>n.s.</i>	***	$\dagger$	***	*	***	<i>n.s.</i>	**	<i>n.s.</i>	
			Across x4 BLs: $\chi^2(3, N = 26) = 5.455, p = .141$, Kendall's $W = .070$		Across x3 BLs: $\chi^2(2, N = 26) = 4.308, p = .116$, Kendall's $W = .083$						
			x1 week post-treatment vs. Final BL: $p = .0156^*$		x1 week post-treatment vs. Final BL: $p = .039^{\dagger}$						
			x4 weeks post-treatment vs. Final BL: $p = .289$		x4 weeks post-treatment vs. Final BL: $p = .008^{**}$						
					x7 weeks post-treatment vs. Final BL: $p = .227$						

Treatment vs. Rep. Exp. Counterpart (where nec.)					Final BLs: $p > .90$ x1 week post-treatment: $p = .549$ x4 weeks post-treatment: $p > .90$ x7 weeks post-treatment: $p = .581$					
RT for Correct Responses										
WEST	***	<i>n.s.</i>	***	*	***	*	***	*	***	<i>n.s.</i>
Follow-up (where nec.)			Across x4 BLs: $\chi^2(3, N = 26) = 6.080, p = .108$, Kendall's $W = .078$ x1 week post-treatment vs. Final BL: $p = .001002^{**}$ x4 weeks post-treatment vs. Final BL: $p = .098$		Across x3 BLs: $\chi^2(2, N = 26) = 4.933, p = .085$, Kendall's $W = .095$ x1 week post-treatment vs. Final BL: $p = .018^{\dagger}$ x4 weeks post-treatment vs. Final BL: $p = .006^{**}$ x7 weeks post-treatment vs. Final BL: $p = .227$		Across x3 BLs: $\chi^2(2, N = 26) = 1.625, p = .444$, Kendall's $W = .031$ x1 week post-treatment vs. Final BL: $p = .009^{**}$ x4 weeks post-treatment vs. Final BL: $p = .003^{**}$ x7 weeks post-treatment vs. Final BL: $p = .002^{**}$			
Treatment vs. Rep. Exp. Counterpart (where nec.)					Final BLs: $p = .646$ x1 week post-treatment: $p = .848$ x4 weeks post-treatment: $p = .794$ x7 weeks post-treatment: $p = .115$					

WEST: $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's):
 $\dagger .025 < p < .05$, $* .01 < p < .025$, $** .001 < p < .01$, $*** p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's):
 $\dagger .01667 < p < .05$, $* .01 < p < .01667$, $** .001 < p < .01$, $*** p < .001$
- Comparing post-treatment outcome against its repeated exposure counterpart (each time-point compared just once):
 $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

Condition	Initial Letter (Orthography)		Repeated Exposure: Initial Letter (Orthography)		Whole Word (Orthography)		Repeated Exposure: Whole Word (Orthography)		Unseen Control: Untreated, Unexposed	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P2										
Initial Naming Accuracy										
WEST	**	**	*	†	***	***	**	n.s.	n.s.	n.s.
Follow-up (where nec.)	Across x4 BLs: $\chi^2(3, N = 24) = 0.243, p = .970$, Kendall's $W = .003$ x1 week post-treatment vs. Final BL: $p = .0156^*$ x4 weeks post-treatment vs. Final BL: $p = .004^{**}$		Across x4 BLs: $\chi^2(3, N = 24) = 0.257, p = .968$, Kendall's $W = .004$ x1 week post-treatment vs. Final BL: $p = .125$ x4 weeks post-treatment vs. Final BL: $p = .289$		Across x3 BLs: $\chi^2(2, N = 24) = 0.222, p = .895$, Kendall's $W = .005$ x1 week post-treatment vs. Final BL: $p = .0001^{***}$ x4 weeks post-treatment vs. Final BL: $p = .065$ x7 weeks post-treatment vs. Final BL: $p = .70$					
Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p > .90$ x1 week post-treatment: $p = .791$ x4 weeks post-treatment: $p = .302$				Final BLs: $p > .90$ x1 week post-treatment: $p = .013^*$ x4 weeks post-treatment: $p = .754$ x7 weeks post-treatment: $p > .90$					
Final Naming Accuracy										
WEST	***	***	*	*	***	***	**	**	†	n.s.
Follow-up (where nec.)	Across x4 BLs: $\chi^2(3, N = 24) = 0.300, p = .960$, Kendall's $W = .004$ x1 week post-treatment vs. Final BL: $p = .002^{**}$ x4 weeks post-treatment vs. Final BL: $p = .002^{**}$		Across x4 BLs: $\chi^2(3, N = 24) = 0.243, p = .970$, Kendall's $W = .003$ x1 week post-treatment vs. Final BL: $p = .031^\dagger$ x4 weeks post-treatment vs. Final BL: $p = .289$		Across x3 BLs: $\chi^2(2, N = 24) = 0.182, p = .913$, Kendall's $W = .004$ x1 week post-treatment vs. Final BL: $p = .00006^{***}$ x4 weeks post-treatment vs. Final BL: $p = .004^{**}$ x7 weeks post-treatment vs. Final BL: $p = .006^{**}$		Across x3 BLs: $\chi^2(2, N = 24) = 0.182, p = .913$, Kendall's $W = .004$ x1 week post-treatment vs. Final BL: $p = .008^{**}$ x4 weeks post-treatment vs. Final BL: $p = .031^\dagger$ x7 weeks post-treatment vs. Final BL: $p = .031^\dagger$			

Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p > .90$ x1 week post-treatment: $p = .549$ x4 weeks post-treatment: $p = .267$	Final BLs: $p > .90$ x1 week post-treatment: $p = .039^*$ x4 weeks post-treatment: $p = .146$ x7 weeks post-treatment: $p = .344$	
RT for Correct Responses			
WEST	*** Across x4 BLs: $\chi^2(3, N = 24) = 0.506, p = .918$, Kendall's $W = .007$ x1 week post-treatment vs. Final BL: $p = .0012^{**}$ x4 weeks post-treatment vs. Final BL: $p = .002^{**}$	** Across x4 BLs: $\chi^2(3, N = 24) = 1.081, p = .782$, Kendall's $W = .015$ x1 week post-treatment vs. Final BL: $p = .006^{**}$ x4 weeks post-treatment vs. Final BL: $p = .041^\dagger$	* Across x4 BLs: $\chi^2(3, N = 24) = 1.081, p = .782$, Kendall's $W = .015$ x1 week post-treatment vs. Final BL: $p = .006^{**}$ x4 weeks post-treatment vs. Final BL: $p = .041^\dagger$
	*** Across x3 BLs: $\chi^2(2, N = 24) = 0.389, p = .823$, Kendall's $W = .008$ x1 week post-treatment vs. Final BL: $p = .0003^{***}$ x4 weeks post-treatment vs. Final BL: $p = .002^{**}$ x7 weeks post-treatment vs. Final BL: $p = .002^{**}$	*** Across x3 BLs: $\chi^2(2, N = 24) = 0.389, p = .823$, Kendall's $W = .008$ x1 week post-treatment vs. Final BL: $p = .0003^{***}$ x4 weeks post-treatment vs. Final BL: $p = .002^{**}$ x7 weeks post-treatment vs. Final BL: $p = .002^{**}$	** Across x3 BLs: $\chi^2(2, N = 24) = 0.056, p = .973$, Kendall's $W = .001$ x1 week post-treatment vs. Final BL: $p = .0096^{**}$ x4 weeks post-treatment vs. Final BL: $p = .047^\dagger$ x7 weeks post-treatment vs. Final BL: $p = .016605^\dagger$
			<i>n.s.</i> <i>n.s.</i>
Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p > .90$ x1 week post-treatment: $p = .711$ x4 weeks post-treatment: $p = .306$	Final BLs: $p = .068^\dagger (> .05)$ x1 week post-treatment: $p = .011^*$ x4 weeks post-treatment: $p = .053^\dagger$ x7 weeks post-treatment: $p = .163$	

WEST: $\dagger .05 < p < .07$, $*$ $.01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's):
 $\dagger .025 < p < .05$, $*$ $.01 < p < .025$, $** .001 < p < .01$, $*** p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's):
 $\dagger .01667 < p < .05$, $*$ $.01 < p < .01667$, $** .001 < p < .01$, $*** p < .001$
- Comparing post-treatment outcome against its repeated exposure counterpart (each time-point compared just once):
 $\dagger .05 < p < .07$, $*$ $.01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

BL: Baseline

Condition	Initial Letter (Orthography)		Repeated Exposure: Initial Letter (Orthography)		Whole Word (Orthography)		Repeated Exposure: Whole Word (Orthography)		Unseen Control: Untreated, Unexposed	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P3										
Initial Naming Accuracy										
WEST	†	n.s.	**	n.s.	***	***	**	n.s.	**	n.s.
Follow-up (where nec.)					Across x4 BLs: $\chi^2(3, N = 25) = 1.059, p = .787$, Kendall's $W = .014$ x1 week post-treatment vs. Final BL: $p = .0002^{***}$ x4 weeks post-treatment vs. Final BL: $p = .039^{\dagger}$					
Treatment vs. Rep. Exp. Counterpart (where nec.)					Final BLs: $p = .125$ x1 week post-treatment: $p = .065^{\dagger}$ x4 weeks post-treatment: $p > .90$					
Final Naming Accuracy										
WEST	*	n.s.	***	n.s.	***	***	**	n.s.	*	n.s.
Follow-up (where nec.)					Across x4 BLs: $\chi^2(3, N = 25) = 0.767, p = .857$, Kendall's $W = .010$ x1 week post-treatment vs. Final BL: $p = .0002^{***}$ x4 weeks post-treatment vs. Final BL: $p = .035^{\dagger}$					
Treatment vs. Rep. Exp. Counterpart (where nec.)					Final BLs: $p = .0703$ x1 week post-treatment: $p = .065^{\dagger}$ x4 weeks post-treatment: $p > .90$					
RT for Correct Responses										
WEST	*	n.s.	**	n.s.	**	**	**	n.s.	*	n.s.

Follow-up (where nec.)		Across x4 BLs: $\chi^2(3, N = 25) = 1.378, p = .711$, Kendall's $W = .018$ x1 week post-treatment vs. Final BL: $p = .0014^{**}$ x4 weeks post-treatment vs. Final BL: $p = .088$	
Treatment vs. Rep. Exp. Counterpart (where nec.)		Final BLs: $p = .248$ x1 week post-treatment: $p = .0196^*$ x4 weeks post-treatment: $p = .943$	

WEST: † $.05 < p < .07$, * $.01 < p < .05$, ** $.001 < p < .01$, *** $p < .001$

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's):
† $.025 < p < .05$, * $.01 < p < .025$, ** $.001 < p < .01$, *** $p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's):
† $.01667 < p < .05$, * $.01 < p < .01667$, ** $.001 < p < .01$, *** $p < .001$
- Comparing post-treatment outcome against its repeated exposure counterpart (each time-point compared just once):
† $.05 < p < .07$, * $.01 < p < .05$, ** $.001 < p < .01$, *** $p < .001$

Condition	Initial Letter (Orthography)		Repeated Exposure: Initial Letter (Orthography)		Whole Word (Orthography)		Repeated Exposure: Whole Word (Orthography)		Unseen Control: Untreated, Unexposed	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P4										
Initial Naming Accuracy										
WEST	**	*	*	n.s.	*	n.s.	n.s.	n.s.	n.s.	†
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 2.000, p = .368$, Kendall's $W = .036$ x1 week post-treatment vs. Final BL: $p = .063$ x4 weeks post-treatment vs. Final BL: $p = .125$ x7 weeks post-treatment vs. Final BL: $p = .0156^*$									
Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p > .90$ x1 week post-treatment: $p = .125$ x4 weeks post-treatment: $p = .687$ x7 weeks post-treatment: $p = .453$									
Final Naming Accuracy										
WEST	***	***	**	n.s.	***	**	*	*	**	*
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 0.500, p = .779$, Kendall's $W = .009$ x1 week post-treatment vs. Final BL: $p = .000061^{***}$ x4 weeks post-treatment vs. Final BL: $p = .0156^*$ x7 weeks post-treatment vs. Final BL: $p = .0002^{***}$				Across x4 BLs: $\chi^2(3, N = 28) = 5.609, p = .132$, Kendall's $W = .067$ x1 week post-treatment vs. Final BL: $p = .021^*$ x4 weeks post-treatment vs. Final BL: $p = .022^*$		Across x4 BLs: $\chi^2(3, N = 28) = 0.600, p = .896$, Kendall's $W = .007$ x1 week post-treatment vs. Final BL: $p = .375$ x4 weeks post-treatment vs. Final BL: $p = .0703$		Across x3 BLs: $\chi^2(2, N = 28) = 0.500, p = .779$, Kendall's $W = .009$ x1 week post-treatment vs. Final BL: $p = .125$ x4 weeks post-treatment vs. Final BL: $p = .0703$ x7 weeks post-treatment vs. Final BL: $p = .039^\dagger$	
Treatment vs. Rep. Exp. Counterpart	Final BLs: $p > .90$ x1 week post-treatment: $p = .0042^{**}$ x4 weeks post-treatment: $p = .549$						Final BLs: $p = .219$ x1 week post-treatment: $p = .004^{**}$ x4 weeks post-treatment: $p = .092$			

(where nec.)

x7 weeks post-treatment: $p = .227$

RT for Correct Responses

WEST	***	**	**	n.s.	***	*	*	**	†
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 0.500, p = .779$, Kendall's $W = .009$ x1 week post-treatment vs. Final BL: $p = .00044^{***}$ x4 weeks post-treatment vs. Final BL: $p = .04995^\dagger$ x7 weeks post-treatment vs. Final BL: $p = .002^{**}$				Across x4 BLs: $\chi^2(3, N = 28) = 3.813, p = .282$, Kendall's $W = .045$ x1 week post-treatment vs. Final BL: $p = .096$ x4 weeks post-treatment vs. Final BL: $p = .0174^*$		Across x4 BLs: $\chi^2(3, N = 28) = 3.562, p = .313$, Kendall's $W = .042$ x1 week post-treatment vs. Final BL: $p = .799$ x4 weeks post-treatment vs. Final BL: $p = .445$	Across x3 BLs: $\chi^2(2, N = 28) = 0.500, p = .779$, Kendall's $W = .009$ x1 week post-treatment vs. Final BL: $p = .0796$ x4 weeks post-treatment vs. Final BL: $p = .04995^\dagger$ x7 weeks post-treatment vs. Final BL: $p = .028^\dagger$	
Treatment vs. Rep. Exp. Counterpart (where nec.)		Final BLs: $p = .317$ x1 week post-treatment: $p = .012^*$ x4 weeks post-treatment: $p = .347$ x7 weeks post-treatment: $p = .407$				Final BLs: $p = .116$ x1 week post-treatment: $p = .173$ x4 weeks post-treatment: $p = .035^*$			

WEST: $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's):
 $\dagger .025 < p < .05$, $* .01 < p < .025$, $** .001 < p < .01$, $*** p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's):
 $\dagger .01667 < p < .05$, $* .01 < p < .01667$, $** .001 < p < .01$, $*** p < .001$
- Comparing post-treatment outcome against its repeated exposure counterpart (each time-point compared just once):
 $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

BL: Baseline

Condition	Initial Letter (Orthography)		Repeated Exposure: Initial Letter (Orthography)		Whole Word (Orthography)		Repeated Exposure: Whole Word (Orthography)		Unseen Control: Untreated, Unexposed	
WEST Test	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC	Trend	ROC
P5										
Initial Naming Accuracy										
WEST	***	***	***	*	***	***	**	*	*	*
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 0.400$, $p = .819$, Kendall's $W = .007$		Across x3 BLs: $\chi^2(2, N = 28) = 0.500$, $p = .779$, Kendall's $W = .009$		Across x4 BLs: $\chi^2(3, N = 28) = 1.000$, $p = .801$, Kendall's $W = .012$		Across x4 BLs: $\chi^2(3, N = 28) = 0.692$, $p = .875$, Kendall's $W = .008$		Across x3 BLs: $\chi^2(2, N = 28) = 1.143$, $p = .565$, Kendall's $W = .020$	
	x1 week post-treatment vs. Final BL: $p = .000145^{***}$		x1 week post-treatment vs. Final BL: $p = .0703$		x1 week post-treatment vs. Final BL: $p = .000015^{***}$		x1 week post-treatment vs. Final BL: $p = .109$		x1 week post-treatment vs. Final BL: $p = .0703$	
	x4 weeks post-treatment vs. Final BL: $p = .00049^{***}$		x4 weeks post-treatment vs. Final BL: $p = .021^\dagger$		x4 weeks post-treatment vs. Final BL: $p = .013^*$		x4 weeks post-treatment vs. Final BL: $p = .219$		x4 weeks post-treatment vs. Final BL: $p = .453$	
	x7 weeks post-treatment vs. Final BL: $p = .0039^*$		x7 weeks post-treatment vs. Final BL: $p = .012^*$						x7 weeks post-treatment vs. Final BL: $p > .90$	
Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p > .90$				Final BLs: $p > .90$					
	x1 week post-treatment: $p = .031^*$				x1 week post-treatment: $p = .004^{**}$					
	x4 weeks post-treatment: $p = .388$				x4 weeks post-treatment: $p = .065^\dagger$					
	x7 weeks post-treatment: $p > .90$									
Final Naming Accuracy										
WEST	***	***	***	***	***	***	**	**	*	**
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 0.286$, $p = .867$, Kendall's $W = .005$		Across x3 BLs: $\chi^2(2, N = 28) = 0.286$, $p = .867$, Kendall's $W = .005$		Across x4 BLs: $\chi^2(3, N = 28) = 0.706$, $p = .872$, Kendall's $W = .008$		Across x4 BLs: $\chi^2(3, N = 28) = 0.462$, $p = .927$, Kendall's $W = .005$		Across x3 BLs: $\chi^2(2, N = 28) = 0.286$, $p = .867$, Kendall's $W = .005$	
	x1 week post-treatment vs. Final BL: $p = .000076^{***}$		x1 week post-treatment vs. Final BL: $p = .002^{**}$		x1 week post-treatment vs. Final BL: $p = .0000001^{***}$		x1 week post-treatment vs. Final BL: $p = .004^{**}$		x1 week post-treatment vs. Final BL: $p = .039^\dagger$	
	x4 weeks post-treatment vs. Final BL: $p = .000275^{***}$		x4 weeks post-treatment vs. Final BL: $p = .004^{**}$		x4 weeks post-treatment vs. Final BL: $p = .004^{**}$		x4 weeks post-treatment vs. Final BL: $p = .219$		x4 weeks post-treatment vs. Final BL: $p = .289$	
	x7 weeks post-treatment vs. Final BL: $p = .0034^{**}$		x7 weeks post-treatment vs. Final BL: $p = .039^\dagger$						x7 weeks post-treatment vs. Final BL: $p > .90$	
Treatment vs. Rep. Exp. Counterpart	Final BLs: $p > .90$				Final BLs: $p > .90$					
	x1 week post-treatment: $p = .143$				x1 week post-treatment: $p = .000977^{***}$					
	x4 weeks post-treatment: $p = .210$				x4 weeks post-treatment: $p = .012^*$					

(where nec.)

x7 weeks post-treatment: $p = .581$

RT for Correct Responses

WEST	***	***	***	***	***	***	***	**	*	**
Follow-up (where nec.)	Across x3 BLs: $\chi^2(2, N = 28) = 0.286$, $p = .867$, Kendall's $W = .005$	Across x3 BLs: $\chi^2(2, N = 28) =$ 0.286 , $p = .867$, Kendall's $W = .005$	Across x4 BLs: $\chi^2(3, N = 28) = 0.739$, p $= .864$, Kendall's $W = .009$	Across x4 BLs: $\chi^2(3, N = 28) = 0.722$, $p = .868$, Kendall's $W = .009$	Across x3 BLs: $\chi^2(2, N = 28) =$ 0.636 , $p = .727$, Kendall's $W = .011$					
	x1 week post-treatment vs. Final BL: $p = .000902^{***}$	x1 week post-treatment vs. Final BL: $p = .002^{**}$	x1 week post-treatment vs. Final BL: p $= .000012^{***}$	x1 week post-treatment vs. Final BL: p $= .006^{**}$	x1 week post-treatment vs. Final BL: p $= .036^{\dagger}$	x1 week post-treatment vs. Final BL: $p = .013^{*}$				
	x4 weeks post-treatment vs. Final BL: $p = .00021^{***}$	x4 weeks post-treatment vs. Final BL: $p = .003^{**}$	x4 weeks post-treatment vs. Final BL: p $= .00497^{**}$	x4 weeks post-treatment vs. Final BL: p $= .036^{\dagger}$		x4 weeks post-treatment vs. Final BL: $p = .066$				
	x7 weeks post-treatment vs. Final BL: $p = .011^{*}$	x7 weeks post-treatment vs. Final BL: $p = .013^{*}$				x7 weeks post-treatment vs. Final BL: $p = .345$				
Treatment vs. Rep. Exp. Counterpart (where nec.)	Final BLs: $p = .109$		Final BLs: $p = .60$							
	x1 week post-treatment: $p = .078^{\dagger}$		x1 week post-treatment: $p = .002^{**}$							
	x4 weeks post-treatment: $p = .181$		x4 weeks post-treatment: $p = .068^{\dagger}$							
	x7 weeks post-treatment: $p = .936$									

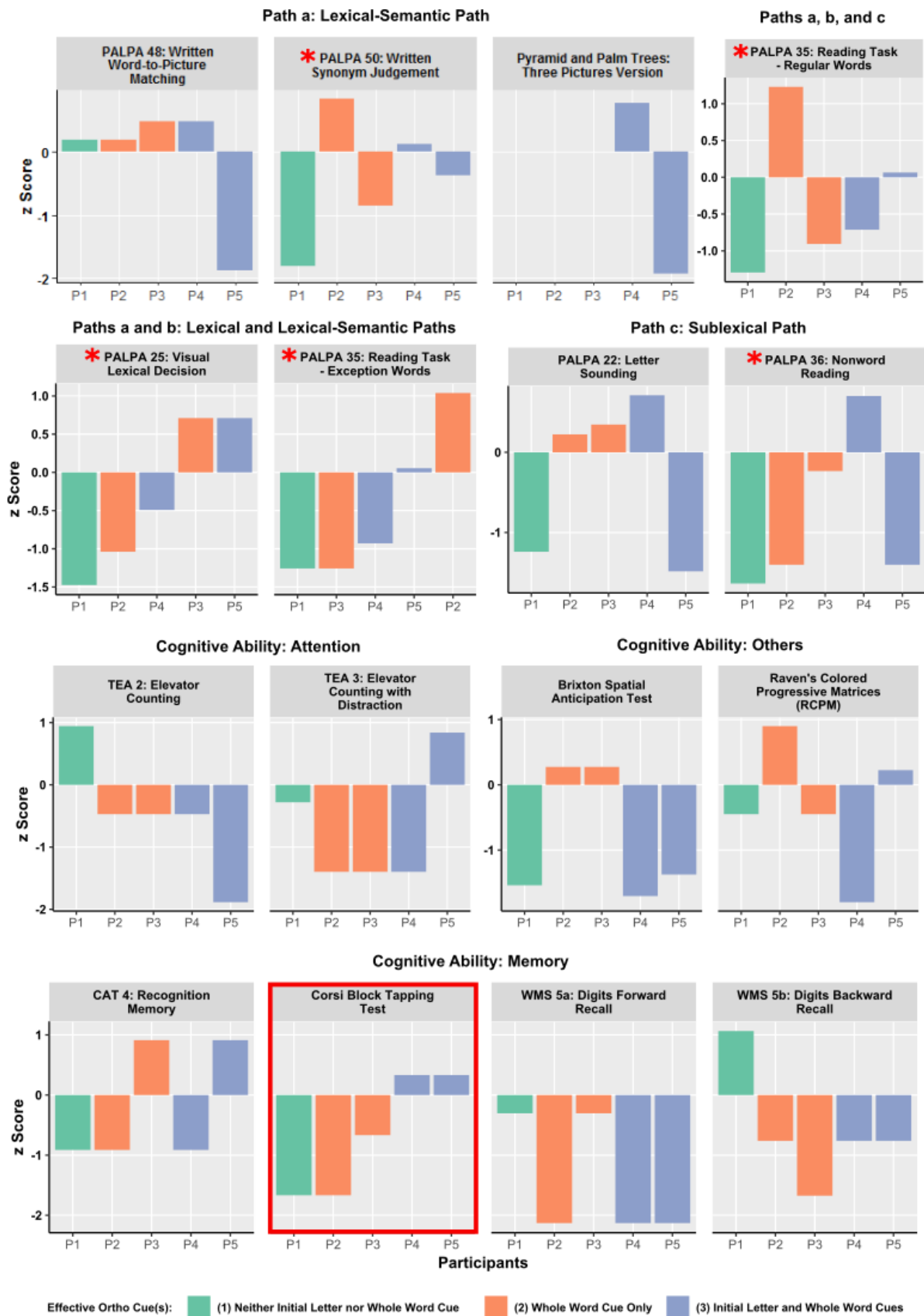
WEST: $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's):
 $\dagger .025 < p < .05$, $* .01 < p < .025$, $** .001 < p < .01$, $*** p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's):
 $\dagger .01667 < p < .05$, $* .01 < p < .01667$, $** .001 < p < .01$, $*** p < .001$
- Comparing post-treatment outcome against its repeated exposure counterpart (each time-point compared just once):
 $\dagger .05 < p < .07$, $* .01 < p < .05$, $** .001 < p < .01$, $*** p < .001$

BL: Baseline

Appendix 6.3

Participants' z-scores on language and cognitive baseline assessments and their groupings according to the therapy cue(s) found effective. Data sorted based on groupings.



Appendix 6.4

Table summarizing the post-therapy changes (Δ) in terms of values for the studies involving the use of orthographic cueing. All post-therapy changes (Δ) are significant, when compared against baseline, according to the original papers unless reported otherwise.

Post-therapy duration	On the Day Therapy Ended	Within 1-week post-therapy	Beyond 1-week post-therapy
Study 1a:	Miceli et al. (1996): Study 1 ('Set 1' in original paper); $n = 1$ (Participant R.B.O.) Facilitation technique: Pre-stimulation with written target words Orthographic unit: Whole word		
Δ in Proportion:	0.58		25 days post-therapy
Δ in Raw Values:	11		0.47 9
Study 1b:	Miceli et al. (1996): Study 2a ('Set 2' in original paper); $n = 1$ (Participant G.M.A.) Facilitation technique: Pre-stimulation with written target words Orthographic unit: Whole word		
Δ in Proportion:	0.91	1-week post-therapy 0.63	2 weeks post-therapy 0.55
Δ in Raw Values:	10	7	6
Study 1c:	Miceli et al. (1996): Study 2b ('Set 1' in original paper); $n = 1$ (Participant G.M.A.) Facilitation technique: Orthographic cueing – Written word simultaneously presented with picture Orthographic unit: Whole word		
Δ in Proportion:	0.91	1-week post-therapy 0.91	4 weeks post-therapy 0.64
Δ in Raw Values:	10	10	7

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to therapy
 Δ in Proportion = Δ in Raw Values / (Total number of items treated minus number of items correctly named at baseline immediately prior to therapy)

Post-therapy duration	On the Day Therapy Ended	Within 1-week post-therapy	Beyond 1-week post-therapy
Study 2a:	Nickels and Best (1996): Study 1 (called ‘ letter cue therapy ’ in the original paper); $n = 1$ (Participant P.A.) Therapy technique: Orthographic task – Given a cue card with four letter choices printed, if unable to correctly name. Read aloud target word printed at the back if still unable to name. Orthographic units: Initial letter and whole word		
Δ in Proportion:	0.56		2 weeks post-therapy
Δ in Raw Values:	9		0.44
			7
Study 2b:	Nickels and Best (1996): Study 2 (called ‘ lexical therapy ’ in the original paper); $n = 1$ (Participant P.A.) Therapy technique: Orthographic cueing – Read aloud the target word, if unable to correctly name Orthographic unit: Whole word		
		1-week post-therapy	2 weeks post-therapy
Δ in Proportion:	0.67	0.40	0.47
Δ in Raw Values:	10	6	7
Study 3a:	Basso et al. (2001): Therapy 1 (‘ Orthographic cue ’ in original paper); $n = 1$ (Participant R.F.) Facilitation technique: Orthographic cueing – Progressive written cues provided if participant did not name the picture correctly within 5 seconds Orthographic unit: Initial letter progressing to the whole word		
		1-week post-therapy	12 weeks post-therapy
Δ in Proportion:	0.90	0.67	0.38
Δ in Raw Values:	19	14	8

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to therapy
 Δ in Proportion = Δ in Raw Values / (Total number of items treated minus number of items correctly named at baseline immediately prior to therapy)

Post-therapy duration	On the Day Therapy Ended	Within 1-week post-therapy	Beyond 1-week post-therapy
Study 3a:	Basso et al. (2001): Therapy 1 ('Orthographic cue' in original paper); $n = 1$ (Participant M.R.) Facilitation technique: Orthographic cueing – Progressive written cues provided if participant did not name the picture correctly within 5 seconds Orthographic unit: Initial letter progressing to the whole word		
Δ in Proportion:	1.00	1-week post-therapy 0.80	5 weeks post-therapy 0.65
Δ in Raw Values:	20	16	13
Study 3b:	Basso et al. (2001): Therapy 2 ('Read cue' or 'Reading' in original paper); $n = 1$ (Participant R.F.) Facilitation technique: Orthographic cueing – Written whole word presented simultaneously with picture Orthographic unit: Whole word		
Δ in Proportion:	0.53	1-week post-therapy 0.14 (<i>n.s.</i>)	12 weeks post-therapy 0.19 (<i>n.s.</i>)
Δ in Raw Values:	11	3 (<i>n.s.</i>)	4 (<i>n.s.</i>)
Study 3b:	Basso et al. (2001): Therapy 2 ('Read cue' or 'Reading' in original paper); $n = 1$ (Participant M.R.) Facilitation technique: Orthographic cueing – Written whole word presented simultaneously with picture Orthographic unit: Whole word		
Δ in Proportion:	0.70	1-week post-therapy 0.30 (<i>n.s.</i>)	5 weeks post-therapy 0.25 (<i>n.s.</i>)
Δ in Raw Values:	14	6 (<i>n.s.</i>)	5 (<i>n.s.</i>)
Study 4:	Leonard, Rochon, and Laird (2004): $n = 1$ (M.H.) Facilitation technique: Orthographic cueing – Progressive written cues provided or self-initiated, if participant did not name the picture correctly. Orthographic unit: Initial letter progressing to the whole word		
Δ in Proportion:	0.80		2 weeks post-therapy 0.1 (<i>n.s.</i> ; likely)
Δ in Raw Values:	8		1 (<i>n.s.</i> ; likely)

Change (Δ) refers to the difference in naming outcomes between post-facilitation time points and the baseline immediately prior to therapy
 Δ in Proportion = Δ in Raw Values / (Total number of items treated minus number of items correctly named at baseline immediately prior to therapy)

Appendix 7

a. Raw numbers (maximum number of ‘opportunities’ for error at each time point) and Poisson main and post hoc analyses of each error type made by P1 across time for initial naming responses.

	(1) Initial Letter Cue (Orthography)						(2) Repeated Exposure: Initial Letter Cue (Orthography)						(3) Whole Word Cue (Orthography)						(4) Repeated Exposure: Whole Word Cue (Orthography)						(5) Unseen Control					
Error Type	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment
Max number of errors at that time point	22	20	18	17	11	12	23	20	20	19	9	14	24	20	19	12	11	15	22	22	20	12	13	9	24	20	18	18	15	17
Semantic (Sem)	1	2	2	8	2	2	2	3	4	4	2	5	1	2	3	2	4	2	4	3	2	1	1	0	2	5	2	5	3	8
	$t(98) = 0.051, p = .959, \eta^2 < 0.001$						$t(103) = 1.42, p = .160, \eta^2 = 0.019$						$t(99) = 1.51, p = .134, \eta^2 = 0.023$						$t(96) = 1.15, p = .254, \eta^2 = 0.014$						$t(110) = 1.96, p = .052, \eta^2 = 0.034^\dagger$					
																									Across x3 BLs: $t(60) = 0.19, p = .847, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(34) = 1.22, p = .231, \eta^2 = 0.042$ x4 wks post-treatment vs. Final BL: $t(31) = 0.51, p = .611, \eta^2 = 0.0085$ x7 wks post-treatment vs. Final BL: $t(33) = 2.08, p = .046, \eta^2 = 0.12^\dagger$					
Sem: Circumlocution, description	1	3	3	1	0	0	2	5	3	6	1	2	1	2	5	2	1	2	2	2	3	1	1	0	6	0	2	4	4	2
	$t(98) = 1.55, p = .125, \eta^2 = 0.024$						$t(103) = 0.64, p = .524, \eta^2 = 0.0040$						$t(99) = 0.062, p = .951, \eta^2 < 0.001$						$t(96) = 0.78, p = .437, \eta^2 = 0.0063$						$t(110) = 0.93, p = .356, \eta^2 = 0.0077$					

Phonological: < 50% similar	1 0 0 1 0 0 $t(98) = 0.77, p = .445, \eta^2 = 0.0060$	0 0 0 1 0 0 $t(103) = 0.53, p = .601, \eta^2 = 0.0027$	0 1 0 0 0 1 $t(99) = 0.36, p = .721, \eta^2 = 0.0013$	0 1 0 0 0 0 $t(96) = 0.72, p = .471, \eta^2 = 0.0054$	0 0 0 0 0 0 —
Phonological: $\geq$ 50% similar	0 0 1 2 0 0 $t(98) = 0.94, p = .349, \eta^2 = 0.0090$	0 1 0 1 0 0 $t(103) = 0.74, p = .459, \eta^2 = 0.0053$	1 0 1 1 0 1 $t(99) = 0.51, p = .614, \eta^2 = 0.0026$	0 1 1 0 1 0 $t(96) = 0.049, p = .961, \eta^2 < 0.001$	0 0 1 1 0 0 $t(110) = 0.15, p = .880, \eta^2 < 0.001$
Semantic/ Phonological	0 0 2 1 0 2 $t(98) = 0.90, p = .371, \eta^2 = 0.0082$	3 1 1 1 2 3 $t(103) = 1.90, p = .060, \eta^2 = 0.034^\dagger$ Across x4 BLs: $t(80) = 0.72, p = .475, \eta^2 = 0.0064$ x1 wk post-treatment vs. Final BL: $t(26) = 0.61, p = .544, \eta^2 = 0.014$ x4 wks post-treatment vs. Final BL: $t(31) = 1.05, p = .303, \eta^2 = 0.034$	0 1 0 3 0 0 $t(99) = 1.55, p = .125, \eta^2 = 0.024$	0 1 4 1 2 1 $t(96) = 0.61, p = .544, \eta^2 = 0.0039$	0 0 0 1 0 0 $t(110) = 1.11, p = .270, \eta^2 = 0.011$
Orthographic	0 0 0 0 0 0 —	0 0 0 0 0 0 —	0 1 0 0 0 0 $t(99) = 0.77, p = .442, \eta^2 = 0.0060$	0 0 0 0 0 0 —	0 0 0 0 0 0 —
Rejection: Reject Incorrect Response(s)	1 1 2 0 0 0 $t(98) = 1.09, p = .279, \eta^2 = 0.012$	0 0 1 2 0 0 $t(103) = 0.91, p = .364, \eta^2 = 0.0080$	0 0 2 1 1 0 $t(99) = 0.51, p = .614, \eta^2 = 0.0026$	0 1 3 1 0 2 $t(96) = 0.45, p = .654, \eta^2 = 0.0021$	$t(110) = 1.94, p = .054, \eta^2 = 0.033^\dagger$ Across x3 BLs: — x1 wk post-treatment vs. Final BL: $t(34) = 1.47, p = .152, \eta^2 = 0.060$ x4 wks post-treatment vs. Final BL: — x7 wks post-treatment vs. Final BL: $t(33) = 1.02, p = .314, \eta^2 = 0.031$
Other errors: Real word; no link to target	0 0 0 0 0 0 —	1 1 0 0 0 0 $t(103) = 0.74, p = .459, \eta^2 = 0.0053$	1 0 0 0 0 0 $t(99) = 0.77, p = .442, \eta^2 = 0.0060$	0 0 1 0 1 0 $t(96) = 0.45, p = .653, \eta^2 = 0.0021$	0 0 0 1 0 0 $t(110) = 1.11, p = .270, \eta^2 = 0.011$
Other errors: Non- word; no link to target	0 0 0 0 0 1 $t(98) = 1.84, p = .068, \eta^2 = 0.033^\dagger$ Across x4 BLs: — x1 wk post-treatment vs. Final BL: — x4 wks post-treatment vs. Final BL: $t(27) = 1.01, p = .323, \eta^2 = 0.036$	0 1 2 0 0 0 $t(103) = 0.91, p = .364, \eta^2 = 0.0080$	1 0 0 0 1 0 $t(99) = 0.36, p = .721, \eta^2 = 0.0013$	0 0 1 0 0 0 $t(96) = 0.72, p = .471, \eta^2 = 0.0054$	1 1 0 0 0 0 $t(110) = 1.27, p = .208, \eta^2 = 0.014$

Other errors: Multiple words; no link to target	0 2 0 0 0 0 $t(98) = 0.77, p = .445, \eta^2 = 0.0060$	0 0 1 0 0 0 $t(103) = 0.53, p = .601, \eta^2 = 0.0027$	1 0 0 0 0 0 $t(99) = 0.77, p = .442, \eta^2 = 0.0060$	1 1 0 1 0 1 $t(96) = 0.64, p = .525, \eta^2 = 0.0042$	0 0 1 0 1 0 $t(110) = 0.15, p = .880, \eta^2 < 0.001$
Miscellaneous errors	1 1 0 0 0 0 $t(98) = 0.77, p = .445, \eta^2 = 0.0060$	0 0 0 0 1 1 $t(103) = 2.74, p = .007, \eta^2 = 0.068^{**}$ Across x4 BLs: — x1 wk post-treatment vs. Final BL: $t(26) = 1.02, p = .317, \eta^2 = 0.039$ x4 wks post-treatment vs. Final BL: $t(31) = 1.02, p = .317, \eta^2 = 0.032$	1 0 0 0 0 0 $t(99) = 0.77, p = .442, \eta^2 = 0.0060$	1 0 1 1 1 1 $t(96) = 1.19, p = .239, \eta^2 = 0.014$	1 0 0 1 0 0 $t(110) = 0.15, p = .880, \eta^2 < 0.001$
No response	17 11 8 4 9 7 $t(98) = 0.99, p = .327, \eta^2 = 0.0098$	15 8 8 4 3 3 $t(103) = 1.12, p = .265, \eta^2 = 0.012$	17 13 8 3 4 9 $t(99) = 1.21, p = .229, \eta^2 = 0.015$	14 12 4 6 6 4 $t(96) = 0.013, p = .990, \eta^2 < 0.001$	14 14 12 3 7 6 $t(110) = 2.46, p = .015, \eta^2 = 0.052^*$ Across x3 BLs: $t(60) = 0.027, p = .978, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(34) = 2.28, p = .029, \eta^2 = 0.13^{\dagger}$ x4 wks post-treatment vs. Final BL: $t(31) = 0.98, p = .336, \eta^2 = 0.030$ x7 wks post-treatment vs. Final BL: $t(33) = 5.91, p < .001, \eta^2 = 0.51^{***}$

$\dagger .05 < p < .07$, * $.01 < p < .05$, ** $.001 < p < .01$, *** $p < .001$

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):

$\dagger .0175 < p < .05$, * $.01 < p < .0175$, ** $.001 < p < .01$, *** $p < .001$

- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):

$\dagger .0125 < p < .05$, * $.01 < p < .0125$, ** $.001 < p < .01$, *** $p < .001$

—: Not applicable for computation as values are zero across the time points

b. Raw numbers (maximum number of ‘opportunities’ for error at each time point) and Poisson main and post hoc analyses of each error type made by P2 across time for initial naming responses.

	(1) Initial Letter Cue (Orthography)						(2) Repeated Exposure: Initial Letter Cue (Orthography)						(3) Whole Word Cue (Orthography)						(4) Repeated Exposure: Whole Word Cue (Orthography)						(5) Unseen Control					
Error Type	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment
Max number of errors at that time point	20	19	20	20	13	11	20	19	20	20	15	16	19	20	20	6	13	14	21	21	20	16	15	15	20	19	21	18	17	19
Semantic (Sem)	3 $t(101) = 0.24, p = .807, \eta^2 < 0.001$	3	1	1	2	0	3	2	1	2	2	5	3	2	1	0	2	3	1	3	1	4	3	3	1	2	5	2	3	5
																			$t(106) = 1.90, p = .060, \eta^2 = 0.033^{\dagger}$ Across x3 BLs: $t(60) = 0.025, p = .980, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(34) = 1.34, p = .190, \eta^2 = 0.05$ x4 wks post-treatment vs. Final BL: $t(33) = 0.99, p = .332, \eta^2 = 0.029$ x7 wks post-treatment vs. Final BL: $t(33) = 0.99, p = .332, \eta^2 = 0.029$						$t(112) = 0.69, p = .491, \eta^2 = 0.0042$					
Sem: Circumlocution, description	0	0	4	2	2	2	0	1	2	0	1	1	0	3	5	1	0	1	1	4	6	2	2	1	0	2	5	3	3	4
	$t(101) = 1.25, p = .215, \eta^2 = 0.015$						$t(108) = 0.58, p = .561, \eta^2 = 0.0031$						$t(90) = 1.04, p = .301, \eta^2 = 0.012$						$t(106) = 0.91, p = .364, \eta^2 = 0.0078$						$t(112) = 0.94, p = .349, \eta^2 = 0.0078$					
Phonological: < 50% similar	0	0	0	0	0	0	1	0	1	1	1	1	0	1	0	0	1	1	2	2	0	0	0	1	0	0	1	0	3	1
	—						$t(108) = 0.58, p = .561, \eta^2 = 0.0031$						$t(90) = 1.11, p = .271, \eta^2 = 0.013$						$t(106) = 1.02, p = .311, \eta^2 = 0.0097$						$t(112) = 1.46, p = .146, \eta^2 = 0.019$					

Phonological: $\geq 50\%$ similar	1 1 1 1 0 1 $t(101) = 0.17, p = .863, \eta^2 < 0.001$	0 2 0 2 1 1 $t(108) = 0.28, p = .782, \eta^2 < 0.001$	0 0 3 0 0 1 $t(90) = 0.45, p = .655, \eta^2 = 0.0022$	3 0 0 0 0 0 $t(106) = 1.49, p = .138, \eta^2 = 0.021$	0 0 0 0 0 0 —
Semantic/ Phonological	1 0 2 1 1 1 $t(101) = 0.58, p = .566, \eta^2 = 0.0033$	0 0 1 0 2 2 $t(108) = 2.63, p = .010, \eta^2 = 0.060^*$ Across x4 BLs: $t(77) = 0.98, p = .328, \eta^2 = 0.012$ x1 wk post-treatment vs. Final BL: $t(33) = 1.43, p = .162, \eta^2 = 0.058$ x4 wks post-treatment vs. Final BL: $t(34) = 1.43, p = .162, \eta^2 = 0.057$	1 3 2 2 1 1 $t(90) = 0.27, p = .788, \eta^2 < 0.001$	1 2 2 1 1 3 $t(106) = 0.47, p = .639, \eta^2 = 0.0021$	0 0 0 1 1 0 $t(112) = 1.49, p = .138, \eta^2 = 0.019$
Orthographic	0 0 0 0 0 0 —	0 0 0 0 0 0 —	0 0 0 0 0 0 —	0 0 0 0 0 0 —	0 0 0 0 0 0 —
Rejection: Reject Incorrect Response(s)	1 1 0 4 2 1 $t(101) = 0.71, p = .481, \eta^2 = 0.0049$	0 2 1 2 0 1 $t(108) = 0.62, p = .535, \eta^2 = 0.0036$	0 0 0 1 1 0 $t(90) = 1.91, p = .060, \eta^2 = 0.039^\dagger$ Across x3 BLs: — x1 wk post-treatment vs. Final BL: $t(24) = 0.96, p = .345, \eta^2 = 0.037$ x4 wks post-treatment vs. Final BL: $t(31) = 0.97, p = .339, \eta^2 = 0.030$ x7 wks post-treatment vs. Final BL: —	1 0 0 3 2 3 $t(106) = 2.89, p = .005, \eta^2 = 0.073^{**}$ Across x3 BLs: $t(60) = 1.21, p = .230, \eta^2 = 0.024$ x1 wk post-treatment vs. Final BL: $t(34) = 1.76, p = .088, \eta^2 = 0.083$ x4 wks post-treatment vs. Final BL: $t(33) = 1.41, p = .167, \eta^2 = 0.057$ x7 wks post-treatment vs. Final BL: $t(33) = 1.76, p = .088, \eta^2 = 0.086$	0 2 1 5 0 0 $t(112) = 0.85, p = .396, \eta^2 = 0.0064$
Other errors: Real word; no link to target	1 0 0 0 0 0 $t(101) = 0.55, p = .586, \eta^2 = 0.0029$	1 0 0 1 1 0 $t(108) = 0.20, p = .845, \eta^2 = 0.0036$	0 0 1 0 0 0 $t(90) = 0.74, p = .460, \eta^2 = 0.0061$	0 1 0 0 0 0 $t(106) = 0.86, p = .394, \eta^2 = 0.0069$	0 0 1 0 1 3 $t(112) = 1.46, p = .146, \eta^2 = 0.019$
Other errors: Non- word; no link to target	0 0 1 0 1 1 $t(101) = 1.79, p = .077, \eta^2 = 0.031$	0 1 0 0 3 0 $t(108) = 2.10, p = .038, \eta^2 = 0.039^*$ Across x4 BLs: $t(77) = 1.00, p = .319, \eta^2 = 0.013$ x1 wk post-treatment vs. Final BL: $t(33) = 1.78, p = .084, \eta^2 = 0.088$ x4 wks post-treatment vs. Final BL: —	1 2 0 0 2 1 $t(90) = 0.72, p = .476, \eta^2 = 0.0057$	1 3 0 2 0 1 $t(106) = 0.014, p = .989, \eta^2 < 0.001$	2 0 0 0 1 0 $t(112) = 0.48, p = .630, \eta^2 = 0.0021$

Other errors: Multiple words; no link to target	0 0 0 0 0 0	0 0 0 0 0 0	0 1 1 0 0 0	0 0 1 0 0 0	0 0 0 0 0 0
	—	—	$t(90) = 1.05, p = .295, \eta^2 = 0.012$	$t(106) = 0.86, p = .394, \eta^2 = 0.0069$	—
Miscellaneous errors	0 0 1 0 0 0	0 1 1 1 0 0	1 0 0 0 0 1	0 0 1 0 0 0	0 0 0 0 0 0
	$t(101) = 0.55, p = .586, \eta^2 = 0.0029$	$t(108) = 1.08, p = .282, \eta^2 = 0.011$	$t(90) = 0.41, p = .681, \eta^2 = 0.0019$	$t(106) = 0.86, p = .394, \eta^2 = 0.0069$	—
No response	13 14 10 11 5 5 $t(101) = 1.09, p = .279, \eta^2 = 0.012$	15 10 13 11 4 5 $t(108) = 2.17, p = .032, \eta^2 = 0.042^*$ Across x4 BLs: $t(77) = 0.20, p = .839, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(33) = 1.81, p = .080, \eta^2 = 0.090$ x4 wks post-treatment vs. Final BL: $t(34) = 1.47, p = .151, \eta^2 = 0.060$	13 8 7 2 6 5 $t(90) = 0.55, p = .583, \eta^2 = 0.0034$	11 6 9 4 7 3 $t(106) = 0.97, p = .336, \eta^2 = 0.0087$	17 13 8 7 5 6 $t(112) = 2.32, p = .022, \eta^2 = 0.046^*$ Across x3 BLs: $t(58) = 1.85, p = .069, \eta^2 = 0.056$ x1 wk post-treatment vs. Final BL: $t(37) = 0.22, p = .828, \eta^2 = 0.0013$ x4 wks post-treatment vs. Final BL: $t(36) = 0.79, p = .437, \eta^2 = 0.017$ x7 wks post-treatment vs. Final BL: $t(38) = 0.49, p = .626, \eta^2 = 0.0063$

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):

† .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001

- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):

† .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001

—: Not applicable for computation as values are zero across the time points

c. Raw numbers (maximum number of ‘opportunities’ for error at each time point) and Poisson main and post hoc analyses of each error type made by P3 across time for initial naming responses.

	(1) Initial Letter Cue (Orthography)						(2) Repeated Exposure: Initial Letter Cue (Orthography)						(3) Whole Word Cue (Orthography)						(4) Repeated Exposure: Whole Word Cue (Orthography)						(5) Unseen Control					
Error Type	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment
Max number of errors at that time point	20	19	20	17	14	17	21	19	20	18	13	15	21	21	21	23	10	15	21	20	20	18	17	14	23	20	20	18	18	17
Semantic (Sem)	1	1	2	1	1	3	1	2	1	2	2	6	2	3	3	2	0	1	1	0	2	2	1	3	1	3	4	3	2	6
	$t(105) = 0.64, p = .523, \eta^2 = 0.0039$						$t(104) = 2.14, p = .035, \eta^2 = 0.042^*$						$t(109) = 1.06, p = .290, \eta^2 = 0.010$						$t(108) = 1.08, p = .282, \eta^2 = 0.011$						$t(114) = 1.06, p = .290, \eta^2 = 0.0098$					
							Across x3 BLs: $t(58) = 0.022, p = .982, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(36) = 0.60, p = .554, \eta^2 = 0.0098$ x4 wks post-treatment vs. Final BL: $t(31) = 0.60, p = .555, \eta^2 = 0.011$ x7 wks post-treatment vs. Final BL: $t(33) = 2.00, p = .054, \eta^2 = 0.108$																							
Sem: Circumlocution, description	0	1	1	1	0	1	0	0	2	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	2
	$t(105) = 0.20, p = .838, \eta^2 < 0.001$						$t(104) = 1.24, p = .219, \eta^2 = 0.014$						$t(109) = 0.53, p = .594, \eta^2 = 0.0026$						$t(108) = 0.62, p = .535, \eta^2 = 0.0036$						$t(114) = 0.64, p = .525, \eta^2 = 0.0036$					
Phonological: < 50% similar	0	0	0	0	3	0	1	0	2	1	0	3	1	3	0	1	2	1	0	1	4	1	1	1	1	2	0	0	1	3

	$t(105) = 1.94, p = .056, \eta^2 = 0.034^\dagger$ Across x3 BLs: — x1 wk post-treatment vs. Final BL: — x4 wks post-treatment vs. Final BL: $t(32) = 1.78, p = .084, \eta^2 = 0.090$ x7 wks post-treatment vs. Final BL: —						$t(104) = 0.73, p = .468, \eta^2 = 0.0051$						$t(109) = 1.01, p = .315, \eta^2 = 0.0093$						$t(108) = 0.20, p = .843, \eta^2 < 0.001$						$t(114) = 0.60, p = .547, \eta^2 = 0.0032$					
Phonological: $\geq 50\%$ similar	0	0	0	0	0	0	1	0	1	2	2	0	1	0	2	1	1	0	1	2	2	2	5	0	3	1	0	1	1	1
	—						$t(104) = 1.15, p = .254, \eta^2 = 0.012$						$t(109) = 0.13, p = .894, \eta^2 < 0.001$						$t(108) = 1.03, p = .303, \eta^2 = 0.0098$						$t(114) = 0.15, p = .882, \eta^2 < 0.001$					
Semantic/ Phonological	1	1	1	0	0	2	3	4	3	4	3	2	1	0	1	3	0	1	2	1	1	1	0	1	1	3	2	2	3	1
	$t(105) = 0.22, p = .829, \eta^2 < 0.001$						$t(104) = 0.35, p = .730, \eta^2 = 0.0012$						$t(109) = 0.34, p = .734, \eta^2 = 0.0011$						$t(108) = 0.62, p = .535, \eta^2 = 0.0036$						$t(114) = 0.30, p = .767, \eta^2 < 0.001$					
Orthographic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	—						—						—						—						—					
Rejection: Reject Incorrect Response(s)	2	1	2	1	0	0	4	3	2	0	1	0	2	3	5	2	1	2	2	3	0	0	0	2	3	2	5	0	1	0
	$t(105) = 1.39, p = .168, \eta^2 = 0.018$						$t(104) = 2.16, p = .033, \eta^2 = 0.043^*$ Across x3 BLs: $t(58) = 0.76, p = .448, \eta^2 = 0.0099$ x1 wk post-treatment vs. Final BL: $t(36) = 1.39, p = .174, \eta^2 = 0.051$ x4 wks post-treatment vs. Final BL: $t(31) = 0.53, p = .598, \eta^2 = 0.0091$ x7 wks post-treatment vs. Final BL: $t(33) = 1.39, p = .174, \eta^2 = 0.055$						$t(109) = 0.23, p = .817, \eta^2 < 0.001$						$t(108) = 0.023, p = .982, \eta^2 < 0.001$						$t(114) = 2.48, p = .015, \eta^2 = 0.051^*$ Across x3 BLs: $t(61) = 0.86, p = .395, \eta^2 = 0.012$ x1 wk post-treatment vs. Final BL: $t(36) = 2.27, p = .029, \eta^2 = 0.125^\dagger$ x4 wks post-treatment vs. Final BL: $t(36) = 1.58, p = .122, \eta^2 = 0.065$ x7 wks post-treatment vs. Final BL: $t(35) = 2.27, p = .029, \eta^2 = 0.129^\dagger$					
Other errors: Real word; no link to target	0	1	3	2	2	1	0	0	0	1	1	0	0	2	0	2	1	4	0	0	1	0	1	1	1	0	0	2	0	1
	$t(105) = 0.64, p = .523, \eta^2 = 0.0039$						$t(104) = 1.62, p = .108, \eta^2 = 0.025$						$t(109) = 2.41, p = .017, \eta^2 = 0.051^*$ Across x4 BLs: $t(84) = 0.036, p = .972, \eta^2 < 0.001$ x1 wk post-treatment vs. Final BL: $t(31) = 0.56, p = .578, \eta^2 = 0.010$ x4 wks post-treatment vs. Final BL: $t(36) = 0.80, p = .428, \eta^2 = 0.018$						$t(108) = 1.48, p = .141, \eta^2 = 0.020$						$t(114) = 1.17, p = .243, \eta^2 = 0.012$					

Other errors: Non-word; no link to target	2 1 3 7 2 6 $t(105) = 2.50, p = .014, \eta^2 = 0.056^*$ Across x3 BLs: $t(57) = 0.49, p = .627, \eta^2 = 0.0042$ x1 wk post-treatment vs. Final BL: $t(35) = 1.29, p = .206, \eta^2 = 0.045$ x4 wks post-treatment vs. Final BL: $t(32) = 0.42, p = .681, \eta^2 = 0.0054$ x7 wks post-treatment vs. Final BL: $t(35) = 1.02, p = .317, \eta^2 = 0.029$	6 1 1 3 2 2 $t(104) = 0.25, p = .801, \eta^2 < 0.001$	3 2 2 4 2 3 $t(109) = 0.83, p = .408, \eta^2 = 0.0063$	5 5 2 4 3 4 $t(108) = 0.24, p = .812, \eta^2 < 0.001$	2 1 1 2 5 2 $t(114) = 1.71, p = .090, \eta^2 = 0.025$
Other errors: Multiple words; no link to target	0 1 0 0 0 0 $t(105) = 0.90, p = .372, \eta^2 = 0.0076$	0 0 0 0 0 0 —	0 1 0 0 0 0 $t(109) = 0.53, p = .594, \eta^2 = 0.0026$	0 0 0 0 0 0 —	0 0 0 0 0 0 —
Miscellaneous errors	0 0 0 1 0 1 $t(105) = 1.57, p = .119, \eta^2 = 0.023$	0 0 0 1 1 2 $t(104) = 2.32, p = .022, \eta^2 = 0.049^*$ Across x3 BLs: — x1 wk post-treatment vs. Final BL: $t(36) = 1.01, p = .321, \eta^2 = 0.027$ x4 wks post-treatment vs. Final BL: $t(31) = 1.01, p = .322, \eta^2 = 0.032$ x7 wks post-treatment vs. Final BL: $t(33) = 1.45, p = .157, \eta^2 = 0.060$	0 0 1 0 0 0 $t(109) = 0.53, p = .594, \eta^2 = 0.0026$	0 0 0 0 2 0 $t(108) = 2.29, p = .024, \eta^2 = 0.046^*$ Across x4 BLs: — x1 wk post-treatment vs. Final BL: $t(33) = 1.43, p = .163, \eta^2 = 0.058$ x4 wks post-treatment vs. Final BL: —	0 0 0 0 0 0 —
No response	14 12 8 4 6 3 $t(105) = 2.41, p = .018, \eta^2 = 0.053^*$ Across x3 BLs: $t(57) = 1.24, p = .218, \eta^2 = 0.026$ x1 wk post-treatment vs. Final BL: $t(35) = 1.11, p = .274, \eta^2 = 0.034$ x4 wks post-treatment vs. Final BL: $t(32) = 0.49, p = .629, \eta^2 = 0.0074$ x7 wks post-treatment vs. Final BL: $t(35) = 1.48, p = .147, \eta^2 = 0.059$	5 9 8 4 1 0 $t(104) = 2.67, p = .009, \eta^2 = 0.064^{**}$ Across x3 BLs: $t(58) = 0.82, p = .417, \eta^2 = 0.011$ x1 wk post-treatment vs. Final BL: $t(36) = 1.08, p = .288, \eta^2 = 0.031$ x4 wks post-treatment vs. Final BL: $t(31) = 2.41, p = .022, \eta^2 = 0.158^\dagger$ x7 wks post-treatment vs. Final BL: $t(33) = 3.06, p = .004, \eta^2 = 0.221^{**}$	10 7 7 8 3 3 $t(109) = 0.99, p = .325, \eta^2 = 0.0089$	9 8 8 8 4 2 $t(108) = 1.79, p = .077, \eta^2 = 0.029$	10 7 8 8 4 1 $t(114) = 1.42, p = .158, \eta^2 = 0.017$

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):

† .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001

- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):

† .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001

—: Not applicable for computation as values are zero across the time points

d. Raw numbers (maximum number of ‘opportunities’ for error at each time point) and Poisson main and post hoc analyses of each error type made by P4 across time for initial naming responses.

	(1) Initial Letter Cue (Orthography)						(2) Repeated Exposure: Initial Letter Cue (Orthography)						(3) Whole Word Cue (Orthography)						(4) Repeated Exposure: Whole Word Cue (Orthography)						(5) Unseen Control					
Error Type	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment
Max number of errors at that time point	27	28	28	23	24	21	28	27	27	27	26	24	27	28	27	26	26	23	27	28	27	27	25	27	27	27	27	25	24	26
Semantic (Sem)	0	2	2	1	1	1	0	1	1	1	0	1	1	3	2	3	0	0	1	2	0	1	0	0	0	1	0	1	1	2
	$t(149) = 0.11, p = .909, \eta^2 < 0.001$						$t(157) = 0.063, p = .950, \eta^2 < 0.001$						$t(155) = 2.03, p = .044, \eta^2 = 0.026^*$ Across x4 BLs: $t(106) = 0.37, p = .713, \eta^2 = 0.0013$ x1 wk post-treatment vs. Final BL: $t(50) = 1.76, p = .084, \eta^2 = 0.058$ x4 wks post-treatment vs. Final BL: $t(47) = 1.76, p = .085, \eta^2 = 0.062$						$t(159) = 1.38, p = .169, \eta^2 = 0.012$						$t(154) = 1.43, p = .155, \eta^2 = 0.013$					
Sem: Circumlocution, description	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$t(149) = 0.90, p = .369, \eta^2 = 0.0054$						—						$t(155) = 0.67, p = .504, \eta^2 = 0.0029$						—						—					
Phonological: < 50% similar	0	1	1	8	2	3	1	2	1	4	2	3	0	4	4	2	5	4	4	0	2	2	4	2	2	1	6	2	6	4
	$t(149) = 3.34, p = .0011, \eta^2 = 0.070^{**}$ Across x3 BLs:						$t(157) = 1.50, p = .135, \eta^2 = 0.014$						$t(155) = 1.52, p = .130, \eta^2 = 0.015$						$t(159) = 0.84, p = .401, \eta^2 = 0.0044$						$t(154) = 0.83, p = .409, \eta^2 = 0.0044$					

	$t(81) = 0.85, p = .398, \eta^2 = 0.0089$ x1 wk post-treatment vs. Final BL: $t(49) = 2.40, p = .020, \eta^2 = 0.10^\dagger$ x4 wks post-treatment vs. Final BL: $t(50) = 0.56, p = .580, \eta^2 = 0.0062$ x7 wks post-treatment vs. Final BL: $t(47) = 0.98, p = .334, \eta^2 = 0.020$				
Phonological: $\geq 50\%$ similar	0 0 1 1 2 6 $t(149) = 2.92, p = .004, \eta^2 = 0.054^{**}$ Across x3 BLs: $t(81) = 1.22, p = .227, \eta^2 = 0.018$ x1 wk post-treatment vs. Final BL: $t(49) = 0.0092, p = .993, \eta^2 < 0.001$ x4 wks post-treatment vs. Final BL: $t(50) = 0.56, p = .580, \eta^2 = 0.0062$ x7 wks post-treatment vs. Final BL: $t(47) = 1.90, p = .063, \eta^2 = 0.071$	3 0 0 3 2 3 $t(157) = 1.62, p = .108, \eta^2 = 0.016$	1 0 2 1 2 3 $t(155) = 1.58, p = .116, \eta^2 = 0.016$	1 0 1 0 1 3 $t(159) = 1.81, p = .073, \eta^2 = 0.020$	2 3 3 2 4 1 $t(154) = 0.11, p = .914, \eta^2 < 0.001$
Semantic/ Phonological	1 0 2 1 2 0 $t(149) = 0.24, p = .808, \eta^2 < 0.001$	0 2 1 0 1 3 $t(157) = 0.46, p = .647, \eta^2 = 0.0013$	0 2 0 0 3 4 $t(155) = 3.09, p = .002, \eta^2 = 0.058^{**}$ Across x4 BLs: $t(106) = 1.40, p = .166, \eta^2 = 0.018$ x1 wk post-treatment vs. Final BL: $t(50) = 1.74, p = .088, \eta^2 = 0.057$ x4 wks post-treatment vs. Final BL: $t(47) = 2.03, p = .048, \eta^2 = 0.081^\dagger$	2 0 0 3 1 1 $t(159) = 0.21, p = .834, \eta^2 < 0.001$	2 0 4 0 0 1 $t(154) = 1.80, p = .074, \eta^2 = 0.021$
Orthographic	0 0 0 0 0 0 —	0 0 0 1 0 0 $t(157) = 1.03, p = .305, \eta^2 = 0.0067$	0 0 0 0 0 0 —	0 0 0 0 0 0 —	1 0 0 0 0 0 $t(154) = 0.96, p = .339, \eta^2 = 0.0059$
Rejection: Reject Incorrect Response(s)	0 0 0 3 1 0 $t(149) = 2.23, p = .027, \eta^2 = 0.032^*$ Across x3 BLs: — x1 wk post-treatment vs. Final BL: $t(49) = 1.74, p = .089, \eta^2 = 0.058$ x4 wks post-treatment vs. Final BL: $t(50) = 0.98, p = .330, \eta^2 = 0.019$	1 1 1 0 2 0 $t(157) = 0.37, p = .708, \eta^2 < 0.001$	1 0 3 1 0 1 $t(155) = 0.77, p = .445, \eta^2 = 0.0038$	0 2 2 1 1 0 $t(159) = 0.82, p = .416, \eta^2 = 0.0042$	1 1 1 0 0 1 $t(154) = 0.92, p = .359, \eta^2 = 0.0055$

	x7 wks post-treatment vs. Final BL: —																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												</

$t(81) = 1.28, p = .203, \eta^2 = 0.020$
x1 wk post-treatment vs. Final BL:
 $t(49) = 2.85, p = .006, \eta^2 = 0.142^{**}$
 x4 wks post-treatment vs. Final BL:
 $t(50) = 0.91, p = .366, \eta^2 = 0.016$
x7 wks post-treatment vs. Final BL:
 $t(47) = 2.07, p = .044, \eta^2 = 0.084^{\dagger}$

$t(106) = 1.32, p = .189, \eta^2 = 0.016$
x1 wk post-treatment vs. Final BL:
 $t(50) = 2.07, p = .044, \eta^2 = 0.079^{\dagger}$
x4 wks post-treatment vs. Final BL:
 $t(47) = 3.35, p = .002, \eta^2 = 0.193^{**}$

$t(79) = 1.55, p = .124, \eta^2 = 0.030$
 x1 wk post-treatment vs. Final BL:
 $t(50) = 0.68, p = .501, \eta^2 = 0.0091$
 x4 wks post-treatment vs. Final BL:
 $t(49) = 0.25, p = .801, \eta^2 = 0.0013$
 x7 wks post-treatment vs. Final BL:
 $t(51) = 0.82, p = .415, \eta^2 = 0.013$

$\dagger .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001$

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):
 $\dagger .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001$
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):
 $\dagger .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001$

—: Not applicable for computation as values are zero across the time points

e. Raw numbers (maximum number of ‘opportunities’ for error at each time point) and Poisson main and post hoc analyses of each error type made by P5 across time for initial naming responses.

	(1) Initial Letter Cue (Orthography)						(2) Repeated Exposure: Initial Letter Cue (Orthography)						(3) Whole Word Cue (Orthography)						(4) Repeated Exposure: Whole Word Cue (Orthography)						(5) Unseen Control					
Error Type	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 1 wk post-treatment	Week 13: 4 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: Baseline 4	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment	Week 1: Baseline 1	Week 2: Baseline 2	Week 3: Baseline 3	Week 7: 1 wk post-treatment	Week 10: 4 wks post-treatment	Week 13: 7 wks post-treatment
Max number of errors at that time point	27	26	26	10	14	18	26	27	26	20	18	17	27	26	26	25	8	15	27	27	27	26	20	22	25	27	25	19	22	24
Semantic (Sem)	5 $t(119) = 0.12, p = .903, \eta^2 < 0.001$	3	4	2	1	3	2	3	4	3	4	1	2	1	3	3	1	2	4	3	1	4	2	5	2	2	1	4	3	3
Sem: Circumlocution, description	0 $t(119) = 1.26, p = .210, \eta^2 = 0.013$	1	2	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Phonological: < 50% similar	1 $t(119) = 0.45, p = .653, \eta^2 = 0.0017$	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	1	2	0	0	0	0
													$t(125) = 2.15, p = .034, \eta^2 = 0.036^*$ Across x4 BLs: — x1 wk post-treatment vs. Final BL: — x4 wks post-treatment vs. Final BL: $t(38) = 1.00, p = .325, \eta^2 = 0.025$						$t(147) = 0.88, p = .379, \eta^2 = 0.0053$	$t(140) = 1.59, p = .113, \eta^2 = 0.018$										
Phonological: ≥ 50% similar	0 $t(119) = 1.95, p = .053, \eta^2 = 0.031^{\dagger}$	0	0	1	1	0	0	0	1	1	1	2	0	1	1	2	0	0	1	0	0	0	2	0	1	1	1	0	1	0
													$t(125) = 0.94, p = .351, \eta^2 = 0.0070$					$t(147) = 1.48, p = .140, \eta^2 = 0.015$					$t(140) = 0.83, p = .408, \eta^2 = 0.0049$							

	Across x3 BLs: — x1 wk post-treatment vs. Final BL: $t(34) = 1.00, p = .325, \eta^2 = 0.028$ x4 wks post-treatment vs. Final BL: $t(38) = 1.00, p = .325, \eta^2 = 0.026$ x7 wks post-treatment vs. Final BL: —																														
Semantic/ Phonological	3	0	2	2	2	2	2	0	2	1	1	2	2	2	1	1	2	2	1	1	1	2	0	1	1	0	0	2			
	$t(119) = 1.38, p = .170, \eta^2 = 0.016$						$t(132) = 0.51, p = .610, \eta^2 = 0.0020$						$t(125) = 0.15, p = .878, \eta^2 < 0.001$						$t(147) = 0.58, p = .561, \eta^2 = 0.0023$						$t(140) = 0.17, p = .866, \eta^2 < 0.001$						
Orthographic	0	0	0	0	0	0	—	0	0	0	0	0	0	0	0	0	0	—	0	0	0	1	0	0	0	0	0	0			
																			$t(147) = 0.62, p = .534, \eta^2 = 0.0026$												
Rejection: Reject Incorrect Response(s)	1	0	0	1	1	2		0	0	0	0	1	0	2	1	2	0	3	0	0	2	0	2	2	0	0	1	1	1	1	
	$t(119) = 2.15, p = .034, \eta^2 = 0.037^*$ Across x3 BLs: $t(77) = 1.21, p = .231, \eta^2 = 0.019$ x1 wk post-treatment vs. Final BL: $t(34) = 1.00, p = .325, \eta^2 = 0.028$ x4 wks post-treatment vs. Final BL: $t(38) = 1.00, p = .325, \eta^2 = 0.026$ x7 wks post-treatment vs. Final BL: $t(42) = 1.43, p = .160, \eta^2 = 0.046$						$t(132) = 1.20, p = .234, \eta^2 = 0.011$						$t(125) = 1.42, p = .157, \eta^2 = 0.016$						$t(147) = 2.11, p = .036, \eta^2 = 0.030^*$ Across x4 BLs: $t(105) = 1.40, p = .163, \eta^2 = 0.018$ x1 wk post-treatment vs. Final BL: $t(44) = 1.41, p = .164, \eta^2 = 0.043$ x4 wks post-treatment vs. Final BL: $t(46) = 1.41, p = .164, \eta^2 = 0.042$						$t(140) = 1.17, p = .244, \eta^2 = 0.0097$						
Other errors: Real word; no link to target	1	1	1	0	2	1		1	1	0	3	4	2	1	2	2	2	0	2	1	2	2	1	4	2	1	2	1	0	0	3
	$t(119) = 0.78, p = .436, \eta^2 = 0.0051$						$t(132) = 2.81, p = .006, \eta^2 = 0.056^{**}$ Across x3 BLs: $t(77) = 0.86, p = .390, \eta^2 = 0.0096$ x1 wk post-treatment vs. Final BL: $t(44) = 1.74, p = .090, \eta^2 = 0.064$ x4 wks post-treatment vs. Final BL: $t(42) = 2.03, p = .049, \eta^2 = 0.089^\dagger$ x7 wks post-treatment vs. Final BL: $t(41) = 1.40, p = .169, \eta^2 = 0.046$						$t(125) = 0.32, p = .751, \eta^2 < 0.001$						$t(147) = 1.68, p = .094, \eta^2 = 0.019$						$t(140) = 0.15, p = .878, \eta^2 = 0.00017$						
Other errors: Non- word; no link to target	0	4	1	0	3	4		2	3	6	2	2	3	6	5	2	4	0	1	3	3	1	2	2	4	6	3	1	5	3	2
	$t(119) = 1.73, p = .087, \eta^2 = 0.024$						$t(132) = 0.18, p = .854, \eta^2 < 0.001$						$t(125) = 1.38, p = .169, \eta^2 = 0.015$						$t(147) = 1.01, p = .312, \eta^2 = 0.0069$						$t(140) = 0.38, p = .707, \eta^2 = 0.0010$						

Other errors: Multiple words; no link to target	0 0 0 0 0 0	0 0 0 0 0 0	0 0 0 0 0 0	0 0 0 0 0 0	0 0 0 0 0 0
	—	—	—	—	—
Miscellaneous errors	14 9 12 3 4 5 $t(119) = 1.32, p = .189, \eta^2 = 0.014$	14 14 11 8 4 5 $t(132) = 1.63, p = .106, \eta^2 = 0.020$	13 12 12 11 3 8 $t(125) = 0.11, p = .916, \eta^2 < 0.001$	12 9 17 15 6 6 $t(147) = 1.75, p = .082, \eta^2 = 0.020$	11 6 17 8 14 11 $t(140) = 0.57, p = .571, \eta^2 = 0.0023$
No response	2 8 4 0 0 1 $t(119) = 2.31, p = .022, \eta^2 = 0.043^*$ Across x3 BLs: $t(77) = 0.68, p = .496, \eta^2 = 0.0060$ x1 wk post-treatment vs. Final BL: $t(34) = 2.05, p = .048, \eta^2 = 0.110^\dagger$ x4 wks post-treatment vs. Final BL: $t(38) = 2.04, p = .048, \eta^2 = 0.099^\dagger$ x7 wks post-treatment vs. Final BL: $t(42) = 1.32, p = .194, \eta^2 = 0.040$	4 5 1 1 1 2 $t(132) = 0.94, p = .346, \eta^2 = 0.0067$	1 1 3 0 0 0 $t(125) = 1.05, p = .297, \eta^2 = 0.0087$	2 7 5 2 1 1 $t(147) = 1.61, p = .109, \eta^2 = 0.017$	$t(140) = 2.52, p = .013, \eta^2 = 0.043^*$ Across x3 BLs: $t(75) = 0.32, p = .753, \eta^2 = 0.0013$ x1 wk post-treatment vs. Final BL: $t(42) = 0.59, p = .559, \eta^2 = 0.0082$ x4 wks post-treatment vs. Final BL: $t(45) = 1.43, p = .159, \eta^2 = 0.044$ x7 wks post-treatment vs. Final BL: $t(47) = 0.028, p = .978, \eta^2 < 0.001$

† .05 < p < .07, * .01 < p < .05, ** .001 < p < .01, *** p < .001

BL: Baseline

- Follow-up tests for conditions with four baselines (after adjusting for Bonferroni's, including the baseline check):
† .0175 < p < .05, * .01 < p < .0175, ** .001 < p < .01, *** p < .001
- Follow-up tests for conditions with three baselines (after adjusting for Bonferroni's, including the baseline check):
† .0125 < p < .05, * .01 < p < .0125, ** .001 < p < .01, *** p < .001

—: Not applicable for computation as values are zero across the time points