Aphasia therapy software: research, development, and implementation

Cuperus, Pauline

DOI:
10.33612/diss.769923247

CHAPTER 5

THE EFFICACY OF ACTION!: A PILOT STUDY OF AN APHASIA THERAPY APP FOCUSING ON VERBS AND SENTENCES

5.1 BACKGROUND

An ever-growing body of research has found that people with aphasia may improve their language abilities through language therapy provided by speech and language therapists (see e.g. Best et al., 2002; Doesborgh et al., 2004; Greenwood et al., 2010; Takizawa et al., 2015). However, despite its effectiveness, people with aphasia may struggle to access appropriate speech and language therapy for an extended period of time, due to reasons such as financial strain and availability of a therapist (see e.g. Nederlandse Vereniging voor Logopedie en Foniatrie, 2019).

In an effort to address this issue, digital aphasia therapy is increasingly explored as an addition to regular speech and language therapy. Digital therapy options that do not heavily rely on the presence of a speech and language therapist may be a way to increase treatment intensity and duration and increase treatment accessibility, which have been previously argued to be contributing factors to the efficacy of aphasia rehabilitation (see e.g. Brady et al., 2016). Promisingly, previous research regarding the efficacy of digital aphasia therapy has been overwhelmingly positive. Literature reviews have highlighted how digital aphasia therapy can complement regular aphasia therapy (Lavoie et al., 2017; Repetto et al., 2020; Zheng et al., 2016), with reported benefits ranging from reduced impairments (e.g., improved word finding, Lavoie et al., 2017) to increased self-reported confidence in functional communication (e.g. Repetto et al., 2020).

It is noteworthy, however, that many of the digital therapy options described in the literature focus mostly on the single-word level, and predominantly on the use of nouns (see e.g. Adrian et al., 2003; Lavoie et al., 2020; Ramsberger & Marie, 2007). Not as much attention has been given to the use of verbs and to the sentence level. This is an important limitation considering that verbs are particularly difficult for people with aphasia (see e.g.

Bastiaanse & Jonkers, 1998). Nevertheless, there are some studies that do describe digital therapy options focusing on the sentence level and/or on the use of verbs in sentences. Furnas and Edmonds (2014), for example, described a computerised version of Verb Network Strengthening Treatment (VneST). VneST aims to improve lexical retrieval by strengthening the connections between verbs and their corresponding thematic roles (see e.g. Edmonds et al., 2009). Thoroughly tested as an offline, therapist-led treatment, the computerised version of VneST was intended for independent use by people with aphasia. Furnas and Edmunds (2014) found that lexical retrieval improved at the word, sentence, and discourse level for two people with chronic aphasia after eight weeks of treatment. Similarly, Thompson et al. (2010) evaluated the efficacy of software that was based on Treatment of Underlying Forms (TUF). TUF is a therapy that focuses on the treatment of complex, non-canonical sentences, aiming for generalisation of treatment effects to less complex sentence structures (see e.g. Thompson & Shapiro, 2005). Thompson et al. (2010) found that their computerised version of TUF produced effects comparable to TUF provided in-person by clinicians for six people with aphasia.

Based on these two pieces of software focusing on verbs and sentences, it seems that digital aphasia therapy could be a worthwhile option for this relatively underexplored therapy focus (see also Chapter 2 of this dissertation for a more extensive overview of the treatment aims of aphasia therapy software). This is particularly relevant as neither Furnas and Edmonds' (2014) nor Thompson et al.'s (2010) software has since become clinically available as far as we are aware (although a different version of VneST is currently available in digital format, see e.g. Tactus Therapy Solutions Ltd., n.d.). Nevertheless, we observed that there was no evidence-based digital aphasia therapy option available that comprehensively targeted verb processing, from the single word level to complex sentences, and taking into account verb inflection, thematic role assignment, and word order.

It is also important to point out that most of the English-language digital aphasia therapy options focus on the American and UK market, (e.g. Constant Therapy and Tactus Therapy; Constant Therapy, n.d.; Tactus Therapy, n.d.), with other varieties (e.g. Australian English) underrepresented. This was also mentioned by some participants in our survey of clinician views of digital aphasia therapy, who commented that it was important for mobile applications to have "appropriate" accents (Cuperus et al., 2022).

Aiming to address these issues, we developed a new aphasia therapy app (Action!) that focuses on the use of verbs in sentence context. Action! consists of nine treatment steps that incorporates the different levels of processing that are involved in sentence production: Activation of concepts, activation of lemmas and lexemes, and grammatical and phonological encoding (see e.g. Bastiaanse & Van Zonneveld, 2004; Levelt, 1989).

Following Step 1, which works on verb retrieval in isolation, in Step 2, the focus is on retrieving lemmas and lexemes which are elicited in response to the concept represented in the stimuli (consisting of an animation), as participants are asked to retrieve the verb form, and produce and insert it in the present continuous form in a sentence. In Steps 3 and 4, participants are required to complete sentences with verbs inflected for the correct tense, which involves retrieving the appropriate phonological representation of the tense morpheme, as well as grammatical encoding. Steps 5 through 7 follow a Mapping Therapy (see e.g. Byng, 1988; Mitchum et al., 2000; Schwartz et al., 1994) approach, with participants being required to assign thematic roles to entities and insert them in the correct position in the sentence frame. Based on the TUF framework (Thompson & Shapiro, 2005), the final two treatment steps target question formation, with the non-canonical word order of what questions requiring syntactic movement in addition to other processes of grammatical and phonological encoding. In this way, Action! provides a relatively comprehensive treatment programme that allows consideration of an individual's needs in terms of impaired processes when determining an appropriate treatment step (see Chapter 4 for a more elaborate description of Action! as well as the theory underlying the treatment).

In addition to its comprehensive and evidence-based nature, Action! was designed to be as user-friendly as possible. Adhering to the principles of user-centred system design (Bannon, 1986), we considered results from earlier research (Swales et al., 2016), and conducted a survey among speech and language therapists (see Chapter 3; Cuperus et al., 2022), whose recommendations (e.g. using short and clear instructions in a location-appropriate accent) guided us during the development process. As previously suggested by Van de Sandt-Koenderman (2011), we also aimed to make the treatment customisable, with functionally relevant items (selected from the SUBTLEX-UK database; Van Heuven et al., 2014), that could be selected for treatment based on individual user needs.

In this chapter, we describe a pilot study in which two people with aphasia trialled the Action! app for a two-week period. The goal of this pilot study was twofold. Firstly, we aimed to collect pilot data that could provide us with an initial indication as to Action!'s efficacy. Secondly, we aimed to gather feedback from people with aphasia who used the app in order to determine Action!'s suitability for independent use by this target group. Both of these research aims are essential to address in order to further develop the app, particularly with a view to making Action! available for wider clinical use (see also previous pilot studies e.g. Cherney et al., 2021; Grechuta et al., 2017; Palmer et al., 2012).

5.2 METHODOLOGY

5.2.1. Ethics

Ethical approval for this study was provided by the Human Research Ethics committee at Macquarie University (reference number: 3774). Participants provided informed written consent using an aphasia-friendly consent form before embarking on the research.

5.2.2. Materials

Background assessment

Before commencement of the treatment study, we conducted a background assessment consisting of the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004), to obtain a general overview of participants' language abilities; and the Verb and Sentence Test (VAST; Bastiaanse et al., 2002) to assess comprehension and production of verbs and sentences and to identify potential targets for treatment content. We administered only a subset of the CAT subtests, as some tests were covered by the VAST (e.g. spoken sentence comprehension), and some were not a focus of this study (e.g., mathematical processing). The Amsterdam Nijmegen Everyday Language Test (ANELT; Blomert et al., 1995) was used to obtain connected speech samples from the participants for illustration purposes. Further analysis of the ANELT samples to investigate treatment effects on functional communication were outside the scope of the current study due to limited available time. The ANELT prompts were translated into English for use in the current study.

Treatment materials

The aphasia treatment in the current study was provided through the Action! app. The Action! app is an aphasia therapy app that focuses on the use of verbs in sentences. The Action! app consists of nine treatment steps (see Table 5.1). It contains animations depicting intransitive verbs (n = 54) and non-reversible transitive verbs (n = 55), as well as animations depicting reversible transitive verbs (n=42) which are not used in the current chapter; see Chapter 4 for more information). These stimuli are treated in different sentence contexts, depending on the treatment step. The animations depict characters performing the selected actions. The animations were created using Vyond software (*Vyond*, n.d.).

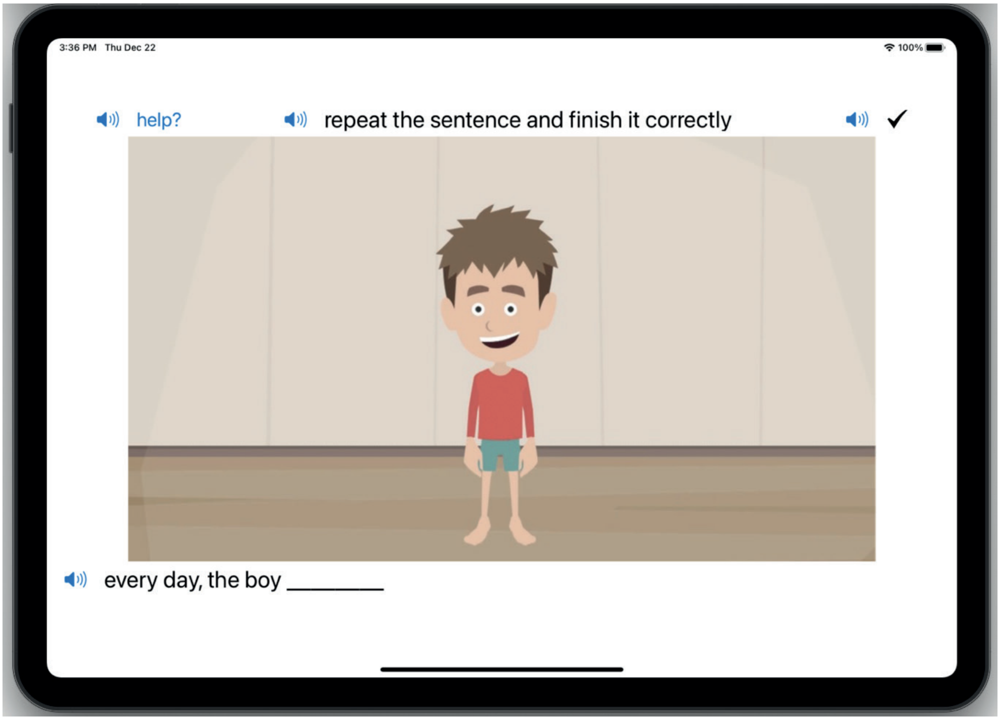
Table 5.1 Overview of the steps included in the Action! programme

Step number	Step name	Short description	Number of items in this step	Cues that can be used in this step
1	Naming actions with verbs in isolation	Participants name actions after watching the animation. No sentence context is provided.	54 intransitive verbs 55 transitive verbs (non-reversible sentences)	• High probability cloze sentence • Initial phoneme cue • Auditory and written presentation of the intended verb
2	Inserting present continuous form in sentences	Participants see a sentence which they need to complete using the present continuous (-ing) form of the intended verb.	54 intransitive verbs 55 transitive verbs (non-reversible sentences)	• High probability cloze sentence • Initial phoneme cue • Auditory and written presentation of the intended verb in sentence context
3	Sentences with verb in habitual present	Participants see a sentence which they need to complete using the habitual present (-s) form of the intended verb.	54 intransitive verbs (27 treated) + 55 transitive verbs (non-reversible sentences; 27 treated)	• High probability cloze sentence • Initial phoneme cue • Auditory and written presentation of the intended verb in sentence context
4	Sentences with verbs in past and future tense	Participants see (non-reversible) sentences without verbs that they need to complete using the correct tense of the verb (past or future tense).	54 intransitive verbs + 55 transitive verbs (non-reversible sentences) All presented in both past and future tense (218 animations in total; 108 animations treated)	• Provision of the verb stem that needs to be conjugated • Selection of the correct tense from two tense forms (past and future) of the verb
5	Thematic roles–- non-reversible sentences	Participants are provided with the constituents of the sentence in a random order and are required to order them correctly	55 transitive verbs (non-reversible sentences)	• Provision of the position and identity of the agent element • Provision of the position and identity of the verb element • Provision of the position and identity of the theme element
6	Thematic roles – expanding verb concepts	The participant is asked to put sentence constituents, according to their thematic roles, in the correct position surrounding the verb.	55 transitive verbs (non-reversible sentences)	• Provision of the position and identity of the subject element • Provision of the position and identity of the object element

7	Thematic roles – reversible sentences	Same as step 5, but with reversible sentences.	21 transitive verbs (reversible sentences), presented twice with different order of actors (42 animations in total)	• Provision of the position and identity of the agent element • Provision of the position and identity of the verb element • Provision of the position and identity of the theme element
8	Who questions	The participant is taken through the steps that are needed to create a who-question. They first need to decide what is the “who” element in the sentence and then drag the sentence elements in the correct position to create a who-question.	21 transitive verbs (reversible sentences), presented twice with different order of actors (42 animations in total)	• Provision of the “who” location • Provision of locations of all sentence constituents
9	What questions	The participant is taken through the steps that are needed to create a what-question. They first need to decide what is the “what” element in the sentence and then drag the sentence elements in the correct position to create a what-question.	55 transitive verbs (non-reversible sentences)	• Provision of the “what” location • Provision of locations of all sentence constituents

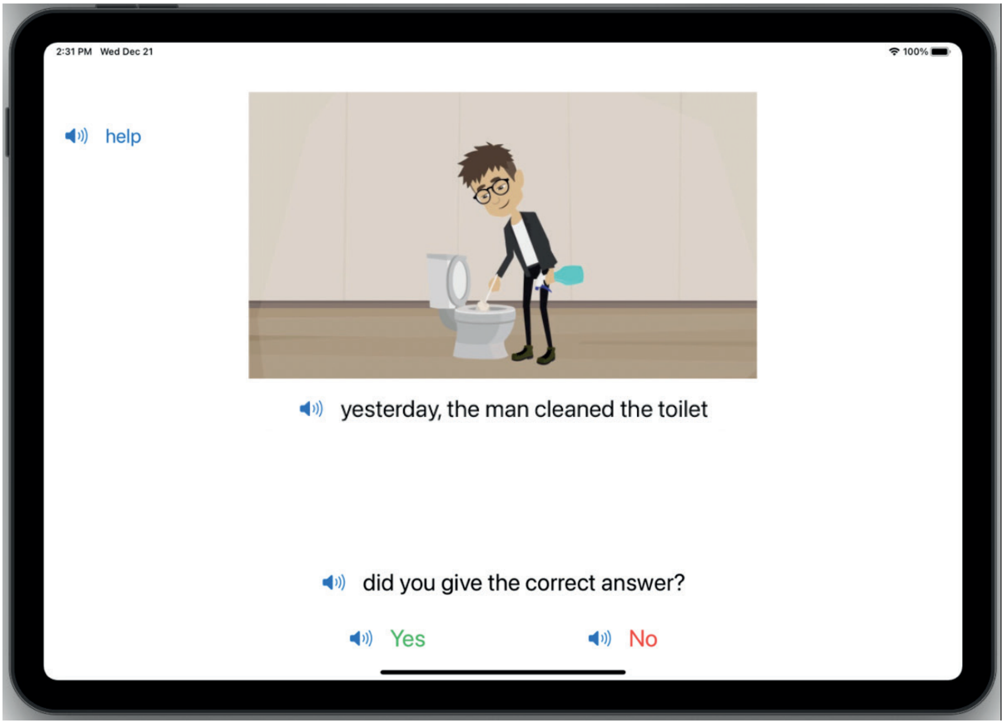
In the current chapter, the focus is on Action! Step 3 (habitual present verb forms) and Step 4 (verbs in past and future tense). Figure 5.1 shows an example of a treatment item in Step 3. The user automatically hears the instructions that are above the animation, as well as the sentence underneath, which they are required to complete in the correct tense. The "help" button allows the user to access the cues described in Table 5.1. Users are instructed to tap the "tick" when they have given an answer. Step 4 has a virtually identical layout and setup to Step 3, although with different sentences, targeting the past tense ("yesterday, the boy _____") and future tense ("tomorrow, the boy _____".

Figure 5.1 Step 3 screenshot for sentence "everyday, the boy smiles"



After tapping the "tick" button, participants are taken to a screen where they are provided with the correct answer in both written and auditory form (see Figure 5.2). Users are asked to indicate whether they gave the correct answer (by tapping "yes") or not (by tapping "no"). They receive feedback in the case of a correct answer (consisting of a game-like positive sound). No feedback is provided for an incorrect answer so as not to discourage the user. The app then moves on automatically to the next item. Chapter 4 of this dissertation provides a more detailed description of the Action! app.

Figure 5.2 Judge accuracy screen for sentence "yesterday, the man cleaned the toilet"



For both participants (JOG and DTR; see participant description in 5.4.2), Action! items were randomly assigned to two sets, one of which was treated while the other set was not. These sets were as similar as possible for pre-treatment accuracy, transitivity, instrumentality, frequency, age of acquisition, and concreteness. We used Discuit (De Kok, 2023) for random assignment of items to sets occurring several times (4 times for JOG; 2 times for DTR), with the best matched assignment being chosen (see Table 5.2 for a Discuit-generated statistical comparison of the treated versus untreated sets on pre-treatment accuracy, transitivity, instrumentality, frequency, age of acquisition, and concreteness, with $p > .05$ indicating that the difference between the two sets was not statistically significant). Item selection occurred after the first two assessment sessions, so that the iPads could be set up and participants could commence treatment immediately after the third assessment session. Therefore, performance at the third assessment session was not considered for determining the treatment items.

Table 5.2 Overview of differences between treated and untreated sets that were part of the treated step on relevant variables

Variable (statistical test)	JOG				DTR			
Pre-treatment accuracy[a]	Measure	Treated set	Untreated set	Statistical difference	Measure	Treated set	Untreated set	Statistical difference
	B1 (lexical retrieval accuracy)	29	30	$c^2(1) = 0.11, p = .74$	B1 (past tense verbs overall accuracy)	21	20	$c^2\ (1) = 0.00, p = 1.0$
	B2 (lexical retrieval accuracy)	27	25	$c^2(1) = 0.00, p = 1.0$	B1 (future tense verbs overall accuracy)	9	9	$c^2\ (1) = 0.00, p = 1.0$
					B2 (past tense verbs overall accuracy)	39	36	$c^2\ (1) = 0.00, p = 1.0$
					B2 (future tense verbs overall accuracy)	27	28	$c^2\ (1) = 0.00, p = 1.0$
Transitivity[a]	Number of transitive verbs	28	27	$c^2\ (1) = 0.00, p = 1.0$	Number of transitive verbs	27	28	$c^2\ (1) = 0.00, p = 1.0$
Instrumentality[a]	Number of instrumental verbs	13	13	$c^2\ (2) = 0.00, p = 1.0$	Number of instrumental verbs	12	14	$c^2\ (2) = 0.16, p = .92$
Frequency[b]	Mean log frequency (*SD*)	1.95 (0.59)	1.99 (0.65)	$c^2\ (1) = 0.19, p = .66$	Mean log frequency (*SD*)	2.00 (0.65)	1.95 (0.59)	$c^2\ (2) = 0.06, p = .81$
Age of acquisition[b]	Mean AoA rating (*SD*)	5.08 (1.16)	5.26 (1.29)	$c^2\ (1) = 0.50, p = .48$	Mean AoA rating (*SD*)	5.28 (1.31)	5.09 (1.13)	$c^2\ (2) = 0.85, p = .36$
Concreteness[b]	Mean concreteness rating (SD)	3.84 (0.59)	3.78 (0.57)	$c^2\ (1) = 0.12, p = .73$	Mean concreteness rating	3.80 (0.56)	3.80 (0.59)	$c^2\ (2) = 0.05, p = .82$

Note: B1 = pre-treatment baseline 1; B2 = pre-treatment baseline 2.
[a] *c^2 with Yates' correction for continuity*
[b] Kruskall-Walis Anova

Outcome measures

Table 5.3 provides an overview of the assessments used to evaluate the effects of the treatment and the background assessment and our rationale for including them in this study.

Table 5.3 Overview of the assessments

Task	Motivation	Background assessment	3 pre-treatment baselines	Post-treatment assessment
Comprehensive Aphasia Test (CAT; Swinburn et al., 2004)	To obtain a general overview of the participant's language abilities and to assess whether participants fulfilled the selection criteria	X		
Verb And Sentence Test (VAST; Bastiaanse et al., 2002)	To assess comprehension and production of verbs and sentences and to determine a potential target for treatment. Also, to assess whether participants fulfilled selection criteria	X		
ANELT (Blomert et al., 1995)	To obtain a connected speech sample in order to describe participants' speech pre-treatment		X	
Treatment-specific outcome measure: Action! step that was treated (treated + untreated items)	To assess whether treatment resulted in step-specific changes		X	X
Control task (nonword repetition)	To evaluate whether there were non-specific effects of treatment/spontaneous recovery		X	X

Treatment-specific outcome measure

As a treatment-specific outcome measure, participants were assessed on the Action! step that they were treated on. In this assessment, participants were shown animations (consisting of treated and untreated intransitive and non-reversible transitive verbs) that were part of Action!. Underneath the animation was a sentence that the participants were asked to complete verbally. The type of sentences that was used here corresponded to the steps that participants were treated on (e.g. "yesterday, the man ______ his teeth" for the past/future step and "every day, the woman ________ popcorn" for the habitual present step).

The participants' responses were coded for accuracy. For Incorrect responses, the type of error made on the first attempted response was coded. For example, in the case of "man is driving no no reversing the car", we coded "driving". The only exception was when participants produced "doing" as a filler, followed by a >3 second-pause, before giving their actual response. "Doing" was then not coded (e.g. for the response "doing the (…), singing" for verb "sings", "singing" would be coded).

Errors were coded as *lexical retrieval* errors (e.g. "push" instead of "pull"), *tense* errors (e.g. "smoked" instead of "will smoke" or "running" instead of "runs"), *phonological* errors (e.g. "cashing" instead of "catching") and/or *no verb produced*. Slightly different versions of these animations (different character/background/direction of the action) had been previously tested with native speakers of Australian English without language impairment (n = 21, mean age = 19.1, 14 female/7 male). These participants had been asked to produce a sentence that matched the action in the animation. For the current study, a response was considered a lexical retrieval error if fewer than 25% of this sample had used the same verb to describe the animation. Errors were coded as phonological errors but not lexical errors if at least 50% of phonemes were identical to the target (e.g. "swinging" for "swimming" was considered a phonological error, not a lexical error). Responses that did not contain a verb were coded as "no verb produced" (e.g. "happy happy birthday" for "celebrating").

Control task

We used nonword repetition as a control task in order to evaluate whether there were any non-specific effects of the treatment, using a list of one-, two-, and three-syllable nonwords from Nickels (1992). Responses on this task were scored as correct or incorrect based on whole nonword accuracy.

Usability

Usability of the Action! app was measured using the System Usability Scale (Brooke, 1996), which had been previously used for this purpose in aphasia research (see e.g. Nef et al., 2018). The System Usability Scale consists of ten statements relating to the usability of software (see Appendix 4.2). These statements touch on different aspects, such as complexity of the system, whether technical help was required, whether the system was easy to learn, and how confident users felt using the system (e.g. "I think that I would like to use this system frequently" and "I needed to learn a lot of things before I could get going with this system"; see Brooke, 1996), which are scored on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The resulting score is converted into a standardised score ranging from 1-100, with 100 being the highest possible usability score. Additionally, the participants were asked about their experiences with Action!, using the usability questionnaire provided in Appendix 4.3. These questions were partially based on previous research by Amaya et al., 2018 and Cherney et al., 2011. The System Usability Scale and the questionnaire were administered during the first post-treatment assessment, immediately after participants had finished the treatment.

5.2.3 Procedure

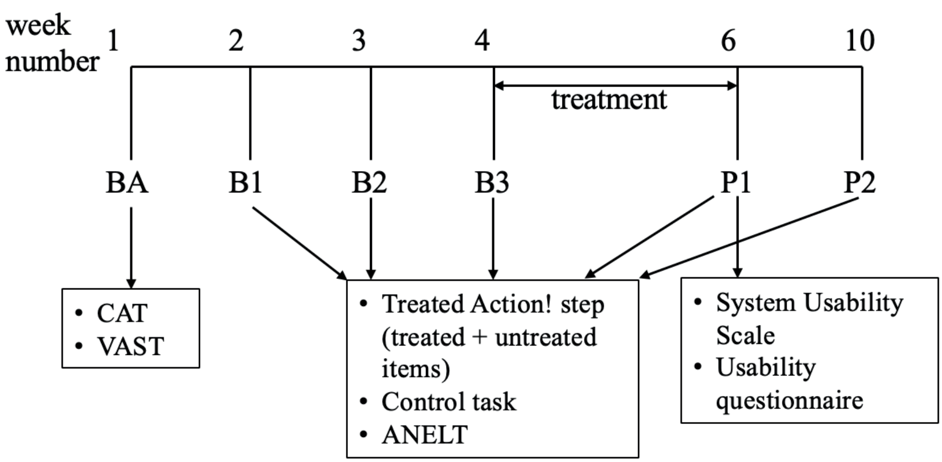
Figure 5.3 visualises the experimental procedure, which lasted a total of ten weeks for each participant. Following administration of the background assessment (BA), pre-treatment baseline assessments were carried out for the treatment-specific task and the control task. These tasks were administered three times over the course of three weeks (pre-treatment baseline assessments B1, B2, B3), thereby establishing an estimate of the level of pre-treatment performance and change over time.

After the background and pre-treatment assessments, participants started the treatment. Treatment lasted for two weeks and consisted of the Action! treatment step that was selected based on participants' CAT and VAST performance (see participant section for rationale for step selection in the current study). Participants were instructed to use the app for at least thirty minutes per day for five days per week, or, at a minimum, to complete the entire treatment step five days per week. Participants were given extensive and aphasia-friendly written instructions on how to use the app and, at the end of the third baseline assessment, practiced several treatment items with the researchers present. Participants could contact the researchers at any time via email or phone during the treatment with questions and were contacted once via phone/email by the research team to check on their progress and to answer questions.

Immediately after finishing the treatment, the outcome measures and the usability measures were administered (post-test 1; P1). The outcome measures were administered once again four weeks after treatment (post-test 2; P2).

All assessments were administered by a speech and language therapist, who was blind to the precise focus of treatment, as well as to which items were treated.

Figure 5.3 The experimental procedure visualised



Note: BA = background assessment; B1 = pre-treatment baseline 1; B2 = pre-treatment baseline 2; B3 = pre-treatment baseline 3; P1 = post-test 1; P2 = post-test 2.

5.2.4 Participants

Two participants were recruited from Macquarie University's Aphasia participant database and from a local aphasia support group. Inclusion criteria were: 1) premorbidly fluent English speaker aged 18-85; 2) in the chronic stage of aphasia (>6 months post-onset); 3) no history of language or cognitive impairments prior to the onset of aphasia; 4) self-reported unimpaired or corrected to normal hearing and vision; 5) impaired verb retrieval and/or sentence construction (as assessed using subtests from the CAT (Swinburn et al., 2004) and the VAST (Bastiaanse et al., 2002); and 6) sufficient iPad skills, which was assessed by self-report prior to participation. Participants were included if they answered "yes" to all questions in Appendix 4.4.

5

Participant 1: JOG

JOG was a 74-year-old ex-accountant. His first language was Italian, but he had lived in Australia since age 9 and premorbidly was a fluent English speaker. JOG had a stroke at age 62 resulting in aphasia. His spontaneous speech was characterised by severe word finding difficulties, agrammatism, long pauses, relatively few lexical verbs, and occasional phonological errors. The three speech samples in Table 5.4 (collected pre-treatment as part of the ANELT (Blomert et al., 1995)) give a picture of JOG's connected speech before the start of the treatment.

Table 5.4 JOG's connected speech pre-treatment (collected as part of the ANELT)

Timepoint	Prompt	JOG's response
B1	You are at the dry cleaners. You have come to pick this up (*a shirt*) and you get it back like this [*present shirt with scorch mark*]. What do you say?	*"All right eh my shirt is eh (.) eh (...) dirty no no eh (..) the eh (.) what's the name of it shirt eh /ɪŋk/ (..) eh now washing washing the shirt dirty because (..) the the eh (..) mark is (.) no (..) wash eh (.) no (...) washing eh (.) shirt no no no eh (...) doing doing doing (..) because (...) no good because eh (.) the (...) the eh (.) finger but no good"*
B2	The kids on the street are playing football in your yard. You have asked them before not to do that. You go outside and speak to the boys. What do you say?	"Eh (..) all right eh (..) after (...) the boy erm (.) every time eh (.) same (.) now (..) eh (.) eh before (..) eh before but now eh (.) no more no more the the eh (.) playing the /bʊt/ /bʊtbɔːl/ / bʊtbɔːl/ because I eh (...) police what's the eh (.) police of eh (..) police yeah"
B3	You are in the chemist and this [present glove] is lying on the floor. What do you say?	"All right (.) pick it up and then eh (.) doing eh (.) the glove eh you (.) glove (.) here and then (.) here here here here pick it up and (..) man is erm (...) all right one glove (...) man chemist chemist here chemist here"

Note: (.) = one-second pause, (..) = two-second pause, (…) = three-second pause

While JOG's comprehension was relatively preserved, the CAT showed impairments on all repetition and spoken language production tasks. Similarly, JOG was impaired on all VAST subtests, with the exception of verb comprehension (see Table 5.5). For sentence comprehension, JOG struggled in particular with object clefts on the VAST subtest.

Because of JOG's difficulties with verb production and finite verbs in particular, we decided Step 3 of Action! was most appropriate, targeting finite, present tense, verb forms using sentence completion. The treatment-specific assessment therefore focused on production of sentences with present tense verbs (e.g. "every day, the woman counts the money").

Due to the severity of JOG's impairment, a reduced subset of 26 one-, two-, and three-syllabic nonwords from the Nickels (1992) list was used as a control task rather than the full set.

Table 5.5 Background assessment results

	Maximum score/ cut-off score (£ cut-off represents impaired performance)	JOG	DTR
Comprehensive Aphasia Test (CAT)			
Comprehension of spoken language			
Comprehension of spoken words	30/25	28	30
Comprehension of spoken paragraphs	4/2	4	4
Comprehension of written language			
Comprehension of written words	30/27	28	30
Comprehension of written sentences	32/23	28	28
Repetition			
Repetition of words	32/29	**24***	30
Repetition of complex words	6/5	**2***	6
Repetition of nonwords	10/5	**0***	6
Repetition of sentences	12/10	**0***	12
Spoken language production			
Naming objects	48/43	**31***	**41***
Naming actions	10/8	**5***	**8***
Word fluency	*NA/13*	***4****	18
Reading aloud			
Reading words	48/45	**34***	48
Reading complex words	6/4	**2***	6
Reading function words	6/3	**6***	6
Reading nonwords	10/6	**0***	8
Verb and sentence test (VAST)			
A – comprehension			
Verb comprehension	40/38	40	39

Sentence comprehension	40/39	**31***	**34***
Grammaticality judgement	40/37	**36***	39
B + C – Production			
Action naming	40/37	**16***	**29***
Filling in finite verbs in sentences	10/8	**0***	**4***
Filling in infinitives in sentences	10/8	**5***	10
Sentence construction	20/16	**6***	19
Sentence anagrams with pictures	20/20	**13***	20
Sentence anagrams without pictures	20/20	**19***	20
Wh-anagrams	18/17	Not adminis-tered	18

Note: Numbers in bold followed by an asterisk (*) represent scores that are below cut-off

Participant 2: DTR

DTR was a 63-year-old former secretary. She was a native English speaker with aphasia following a stroke 6 years prior to this study. DTR's spontaneous speech showed mild word-finding difficulties with frequent fillers and pauses (e.g. "eh"). The three samples in Table 5.6, that were collected pre-treatment as part of the ANELT, provide an illustration of her connected speech.

Table 5.6 DTR's connected speech pre-treatment (collected as part of the ANELT)

Timepoint	Prompt	DTR's response
B1	You have an appointment with the doctor. Something else has come up. You call up and what do you say?	"I would like to (.) eh rebook the eh (..) the date"
B2	You see your neighbour walking by. You want to ask him/her to come to visit some time. What do you say?	*"Eh (..) do you (.) want (.) to come in (..) eh (.) and have a eh of (.) coffee or tea (..) and a chat (...) sometime"*
B3	You have just moved in next door to me. You would like to meet me. You ring my doorbell and say ...	"Erm I'm your nextdoor neighbour (..) eh (.) and my name (.) and (...) would you like to (..) to (.) come in for a cuppa (.) sometime (..) that's it"

Note: (.) = one-second pause, (..) = two-second pause, (…) = three-second pause

The CAT and the VAST confirmed this impression, with some relatively mild impairments on object naming (CAT) and sentence comprehension (struggling mostly with object cleft sentences), action naming, and filling in finite verbs in sentences (using non-finite rather than finite verbs; see Table 5.5).

To determine the most appropriate treatment target, DTR was initially assessed on three Action! treatment steps: Inserting present continuous forms in sentences, sentences with verbs in habitual present, and sentences with verbs in past and future tense (see

Chapter 4 for more information). It was decided that Action! Step 4 (sentences with verbs in past and future tense, e.g. "yesterday, the woman counted the money" or "tomorrow, the woman will count the money; overall accuracy 27.6% at B1) was most suitable for treatment, as this was the step where DTR showed the highest number of errors, with relatively little room for improvement on earlier steps.

5.2.5 Analysis

In these case studies, we compared participants' performance after the treatment to their own performance pre-treatment to establish whether the treatment had effected a change. As advocated by Howard et al. (2015), the scores on treated and untreated items in the Action! treatment-specific assessment pre- and post-therapy, and the control tasks were analysed using WEighted STatistics (WEST). We used the WEST – Rate Of Change (WEST-ROC) to determine whether there was greater change during the treatment phase compared to the no-treatment phase, as well as WEST-TREND to see whether across the study as a whole performance improved. A positive trend as well as a significant difference in the rate of change across phases would allow us to conclude that there had been a positive effect of treatment on performance. The WEST-ROC weights that were used for both participants were: B1: 2; B2: -1; B3: -4; P1: 3; P2: 0 and the WEST-TREND weights that were used were: B1: -2; B2: -1; B3: 0; P1: 1; P2: 2. One sample t-tests (one-tailed) were used to compare the weighted coefficients to the null hypothesis of no change. Two-sample t-tests (two-tailed) were used to compare coefficients across treated and untreated sets. McNemar tests (two-tailed) were used to determine whether scores on the final post-treatment assessment were significantly different from those at the highest-scoring baseline, thereby indicating maintenance effects. The treatment-specific assessment was initially analysed using overall accuracy (items correct for both lexical retrieval and tense), before analysing these two measures separately (correct lexical retrieval regardless of tense, and correct tense regardless of lexical retrieval)

Participants' responses to the usability questionnaire (Appendix 4.2) were transcribed and briefly summarised. Scores on the System Usability Scale were calculated as described by Brooke (1996).

5.3 RESULTS

5.3.1 Efficacy

JOG

Figure 5.4 shows JOG's performance on the treatment-specific assessment across the treatment study. Prior to treatment, JOG's responses on the treatment-specific assessment were consistently at floor. This was due to JOG failing to mark any verb correctly for tense. When looking at his overall accuracy (correct lexical retrieval + correct tense), JOG showed significant improvement over the course of the study, for treated ($t_{WEST\text{-}TREND}(54)$ = 7.61, $p_{WEST\text{-}TREND}$ < .001) and untreated ($t_{WEST\text{-}TREND}(53)$ = 3.26, $p_{WEST\text{-}TREND}$ < .001) verbs. These improvements were treatment-related with greater improvement during the treatment than no treatment phases (treated: $t_{WEST\text{-}ROC}(54)$ = 9.72, $p_{WEST\text{-}ROC}$ < .001; untreated: $t_{WEST\text{-}ROC}(53)$ = 3.26, $p_{WEST\text{-}ROC}$ < .001). Treated verbs were found to have improved significantly more than untreated verbs, both in terms of the change across the phases (WEST-ROC: t = 0.71, p < .001), as well as for the overall trend for improvement (WEST-TREND: t = 0.76, p < .001).

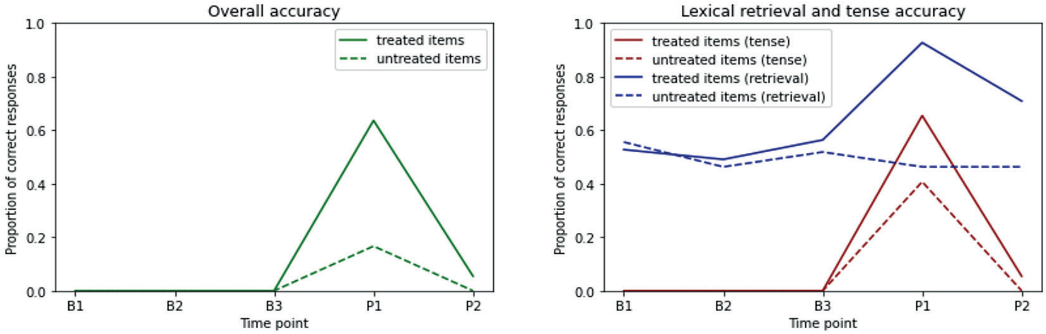
JOG's lexical retrieval accuracy improved significantly over the course of the study for treated items ($t_{WEST\text{-}TREND}(54)$ = 4.68, $p_{WEST\text{-}TREND}$ < .001), which was a result of the treatment as indicated by greater improvement during the treatment than no treatment phases ($t_{WEST\text{-}ROC}(54)$ = 3.55, $p_{WEST\text{-}ROC}$ < .001). This effect was maintained at P2 compared with the highest baseline B3 (McNemar's p < .001). Lexical retrieval scores did not improve significantly for untreated items ($t_{WEST\text{-}TREND}(53)$ = -1.49, $p_{WEST\text{-}TREND}$ = .93), with no difference in the change across treatment and no treatment phases ($t_{WEST\text{-}ROC}(53)$ = -0.15, $p_{WEST\text{-}ROC}$ = .56). Treated verbs improved significantly more than untreated verbs, both in terms of the change across the phases (WEST-ROC: t = 1.03; p = .003), as well as for the overall trend for improvement (WEST-TREND: t = 1.20; p < .001).

There was significant improvement in JOG's inflection of verbs with the correct tense for treated items ($t_{WEST\text{-}TREND}(54)$ = 7.88, $p_{WEST\text{-}TREND}$ < .001), that was related to the treatment ($t_{WEST\text{-}ROC}(54)$ = 10.11, $p_{WEST\text{-}ROC}$ < .001). Similar results were observed for untreated items ($t_{WEST\text{-}TREND}(53)$ = 2.01, $p_{WEST\text{-}TREND}$ < .001; $t_{WEST\text{-}ROC}(53)$ = 6.04, $p_{WEST\text{-}ROC}$ < .001). At P2, however, performance had dropped back close to floor and neither treated nor untreated items were significantly higher than that at the final baseline B3 for tense inflection (treated: McNemar's p = .13; untreated: McNemar's p = 1.0), indicating no significant maintenance of treatment effects. Inflection of treated verbs improved significantly more than untreated verbs, both in terms of the change across the phases (WEST-ROC: t = 6.04, p < .001), as well as for the overall trend for improvement (WEST-TREND: t = 6.04; p < .001).

JOG made phonological errors and "no verb produced" errors across the assessments, with some variability between sessions (e.g. relatively lower phonological error rates and higher "no verb produced" at B2 compared to P1; see Table 5.7). These metrics were not further explored.

JOG's score on the control task was consistently off floor and below ceiling, with accuracy varying between 38.5% and 50.0% pre-treatment (see Table 5.7). JOG did not show significant change on the nonword repetition control task over the course of the study and similarly did not show a significant difference between the amount of change across the two phases ($t_{WEST\text{-}TREND}(25)$ = 0.82, $p_{WEST\text{-}TREND}$ = .21; $t_{WEST\text{-}ROC}(25)$ = -1.47, $p_{WEST\text{-}ROC}$ = .92).

Figure 5.4 JOG's results on the treatment-specific assessment for treated (n = 55) and untreated (n = 54) items, looking at overall accuracy and lexical retrieval and tense separately



Note: BA = background assessment; B1 = pre-treatment baseline 1; B2 = pre-treatment baseline 2; B3 = pre-treatment baseline 3; P1 = post-test 1; P2 = post-test 2.

DTR

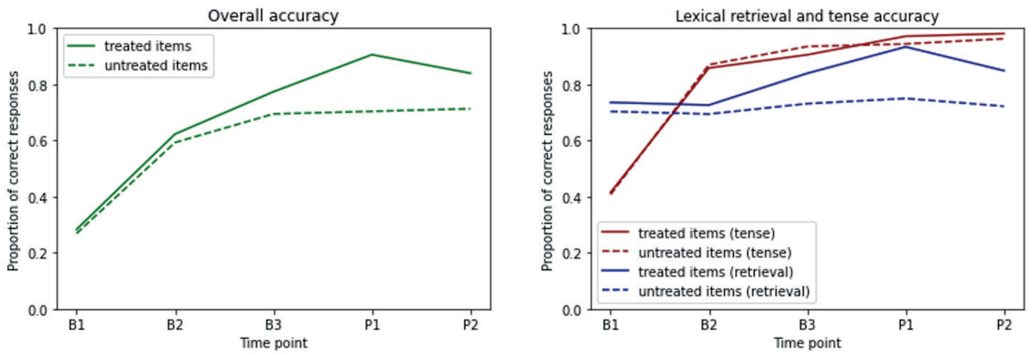
DTR's results for the treatment-specific assessment are visualised in Figure 5.5. It was clear that she showed considerable improvement across the baseline period. While DTR showed improvement in overall accuracy of treated items over the course of the study ($t_{WEST\text{-}TREND}(105) = 10.65$, $p_{WEST\text{-}TREND} < .001$), her rate of improvement during the treatment phase was not significantly different from the rate of change during the no treatment phases ($t_{WEST\text{-}ROC}(105) = -2.14$, $p_{WEST\text{-}ROC} = .98$). DTR also showed significant improvement in overall accuracy of untreated items across all testing points ($t_{WEST\text{-}TREND}(107) = 8.20$, $p_{WEST\text{-}TREND} < .001$). For untreated items, the change as a result of the treatment was found to not be significantly different from the change during the pre-treatment assessments ($t_{WEST\text{-}ROC}(107) = -3.65$, $p_{WEST\text{-}ROC} = 1.00$). The difference in change across the phases for the treated and untreated sets did not differ significantly (WEST-ROC: $t = 0.34$, $p = .15$), but the overall trend for improvement across the study did (WEST-TREND: $t = 0.23$, $p = .01$).

Following visual inspection (see Figure 5.5), performance on tense separately was not statistically analysed as DTR's large improvement during the baseline and close to ceiling performance at B3, made the possibility of any treatment-related effects highly unlikely. Statistical analysis of lexical retrieval revealed that DTR showed an overall significant improvement on lexical retrieval for treated items ($t_{WEST\text{-}TREND}(105) = 3.69$, $p_{WEST\text{-}TREND} < .001$), but her rate of improvement was not significantly different from the rate of change during the baseline ($t_{WEST\text{-}ROC}(105) = 1.16$, $p_{WEST\text{-}ROC} = .12$). For untreated items, DTR did not show overall improvement ($t_{WEST\text{-}TREND}(107) = 1.02$, $p_{WEST\text{-}TREND} = .15$) and the rate of change during the treatment was found to be similar to that across the baseline ($t_{WEST\text{-}ROC}(107) = 0.24$, $p_{WEST\text{-}ROC} = .41$). The difference in change across the phases for the treated and untreated sets did not differ significantly (WEST-ROC: $t = 0.79$, $p = .25$), but the overall trend for improvement across the study did (WEST-TREND: $t = 0.77$, $p = .01$).

5

DTR scored consistently off floor and below ceiling on the 65-item nonword repetition task from Nickels (1992), with accuracy ranging between 78.5% and 83.1% on the pre-treatment baselines (see Table 5.7). DTR did not show significant change on the nonword repetition control task over the course of the study and also did not show a significant difference between the amount of change across the two phases ($t_{WEST\text{-}TREND}(64) = 0.50$, $p_{WEST\text{-}TREND} = .31$; $t_{WEST\text{-}ROC}(64) = -0.85$, $p_{WEST\text{-}ROC} = .80$).

Figure 5.5 DTR's results on the treatment-specific assessment for treated (n = 106) and untreated (n = 108) items, looking at overall accuracy and lexical retrieval and tense separately



Note: BA = background assessment; B1 = pre-treatment baseline 1; B2 = pre-treatment baseline 2; B3 = pre-treatment baseline 3; P1 = post-test 1; P2 = post-test 2.

Table 5.7 JOG and DTR's pre- and post-assessment results on the treated ACTION! step

	JOG						DTR					
	N =	B1	B2	B3	P1	P2	N =	B1	B2	B3	P1	P2
Treatment-specific task (habitual present for JOG, past/future for DTR)												
Overall accuracy (correct verb + correct tense (%))												
Treated	55	0.0	0.0	0.0	63.6	5.4	106	28.3	62.3	77.4	90.6	84.0
Untreated	54	0.0	0.0	0.0	16.7	0.0	108	26.9	59.2	69.4	70.4	71.3
Tense errors (%)												
Treated	55	100.0	100.0	100.0	34.5	94.6	106	58.5	14.2	9.4	2.8	1.9
Untreated	54	100.0	100.0	100.0	59.3	100.0	108	59.3	13.0	6.5	5.6	3.8
Lexical retrieval errors (%)												
Treated	55	47.3	50.9	43.6	7.3	29.1	106	26.4	27.4	16.0	6.7	15.1
Untreated	54	44.4	53.7	48.1	53.7	53.7	108	28.7	30.6	26.9	25.0	27.8
Phonological errors (%)												
Treated	55	10.9	10.9	9.1	5.5	12.7	106	0.0	0.0	0.0	0.0	0.0
Untreated	54	7.4	11.1	3.7	3.7	1.8	108	0.0	0.0	0.0	0.0	0.0
No verb produced (%)												
Treated	55	10.9	5.5	18.2	3.6	5.5	106	0.0	0.0	0.0	0.0	0.0
Untreated	54	13.0	9.3	20.4	5.6	13.0	108	0.0	0.0	0.0	0.0	0.0
Control task – Nonword Repetition												
Correctly repeated nonwords (%)												
	26	38.5	38.5	50.0	34.6	50.0	65	78.5	83.1	81.5	78.5	84.6

Note: BA = background assessment; B1 = pre-treatment baseline 1; B2 = pre-treatment baseline 2; B3 = pre-treatment baseline 3; P1 = post-test 1; P2 = post-test 2.

5.4.2 App usage and usability

JOG

The Action! data show that JOG used the app on thirteen days (with a mean daily usage time of 1.02 hours). He practiced 764 times with individual treatment items, seeing each item 13.9 times on average. The mean time per item was 52.2 seconds (excluding the time needed for judging accuracy). JOG used cues for 39.6% of his attempts, using mostly the "first sound" cue (300 times), as well as the "repeat cue" (3 times) and the "Cloze sentence cue" (3 times). He used multiple cues for some items. JOG indicated that he gave the correct response for 73.3% of his attempts.

In his post-treatment interview, JOG was positive about Action!, stating that it was "easy" to use the app. He noted that at one point he had had some issues logging in, which he had solved with help from his brother. JOG said that he made some mistakes the first two or three times he encountered an item, but generally found the treatment content easy after this (although there were some items that he reported consistently struggling with, particularly "decorates"). When asked whether anything should be changed about the app, JOG responded that the treatment could be "harder". Still, he firmly responded "yes" when asked whether he would recommend Action! to other people with aphasia. JOG's positive experience was further reflected by his rating of 97.5 on the System Usability Scale, on which he "strongly agreed" with all positive statements and "strongly disagreed" with all negative statements, with the exception of "I think that I would like to use this system frequently" (to which he "agreed").

DTR

The data collected by Action! shows that DTR used the app on five different days (averaging 52.5 minutes/day), practicing a total of 721 treatment items over the course of the two weeks that she used the app. She viewed each item an average of 6.8 times in past tense and 6.8 times in future tense. The average time spent on each item was 13.45 seconds (excluding the time needed for judging accuracy following a response). DTR did not use any cues during the treatment and judged her own responses as accurate for 91.3% of attempts.

When discussing the app with DTR post-treatment, she was mostly positive about Action!. DTR liked how the app looked and found the buttons easy to use. She noted that the instructions had been helpful and that she was able to "navigate" herself, saying that it was easy to learn how to use the app. DTR did report some app crashes, saying that the app "stopped a few times and went back to the beginning" (this was later confirmed by the app data and found to be due to an error in the app code). While she found the treatment content "very simple", DTR said she would recommend Action! to other people with aphasia. She rated Action! 95/100 on the System Usability Scale, "strongly agreeing" with most positive statements and "strongly disagreeing" with all negative statements. The only two exceptions were the statements "I think that I would like to use this system frequently" and "I would imagine that most people would learn to use this system very quickly", which DTR "agreed" with.

5.5 DISCUSSION

The goal of the current study was to pilot the newly developed Action! app with two people with aphasia, looking at both efficacy and usability. We will first discuss the efficacy results before examining the usability data and feedback and implications for future development of Action!.

5.5.1 Participant results

Participant JOG, a man with relatively good verb comprehension but severe impairment in verb and sentence production, was treated on the use of habitual present tense. He showed significant improvements as a result of the treatment, with improved lexical retrieval of treated verbs and improved inflection of these verbs for present tense, with improvements on tense generalising to untreated items. However, at four-weeks post-treatment, these results were only maintained for lexical retrieval of treated items, with performance on tense returning to (close to) baseline levels. As JOG's control task scores remained stable throughout the treatment, we can be confident that his results were not due to general recovery or non-specific effects of treatment but were in fact due to the Action! treatment.

Considering that JOG was twelve years post-onset at the time of the study and only used the app for two weeks, his substantial improvement at the first post-treatment assessment was impressive. It is particularly noteworthy that his use of present tense generalised to untreated items. JOG's improved lexical retrieval of treated items is in line with previous research, with improvement of treated items frequently reported in the literature (see e.g. Carragher et al., 2013; Maul et al., 2014; Wambaugh et al., 2014). Improved retrieval of untreated verbs is less frequently reported in the literature (see e.g. De Aguiar et al. (2016) and Webster & Whitworth (2012) for a meta-analysis/review), but has been found to be more likely in people who received morphological training (see e.g. De Aguiar et al., 2016; Links et al., 2010). While JOG did not improve on lexical retrieval of untreated items, he did significantly improve on the overall accuracy (correct lexical retrieval + correct tense) of untreated items, which was mostly due to an improvement in his use of present tense inflection. JOG's results are therefore in line with previous research on treatment of finite verbs (e.g. Links et al., 2010), which did not look at lexical retrieval and accurate tense separately but rather at the overall accuracy. It has been previously argued that generalisation of grammatical information (such as tense) is more likely to occur than generalisation of lexical information as grammatical information is shared across verbs, while lexical information is unique to each verb (see e.g. De Aguiar et al., 2016; Roelofs et al., 1998). JOG's generalisation pattern, with tense improvements generalising to untreated verbs while lexical retrieval did not, supports this theory.

There are several potential factors that may have contributed to the lack of maintenance of the treatment effects on tense inflection for JOG four weeks after the end of treatment. Firstly, it may be that the treatment dose and/or duration that JOG received was simply not sufficient to induce sustained treatment effects, a possibility that is supported by Links et al.'s (2010) results. Links et al. (2010) evaluated the Dutch ACTIE! Programme (on which Action! is partially based; see Chapter 4) in eleven people with aphasia and found that improvement of untrained finite verbs was maintained at three months post-treatment. However, their participants received a total of twelve weeks of treatment (totalling 18 hours), which is a substantially higher dose and longer duration than JOG's treatment of approximately 13 hours over a two-week period. A second potential contributor to JOG's relatively low assessment scores at P2 may have been that just before this assessment session, JOG had attended a two-hour physiotherapy appointment. While difficult to prove, both researchers present at the final assessment session suspected that this intense appointment had had a detrimental effect on JOG's energy and concentration levels during the assessment session, which may have negatively affected his performance on the assessment tasks. Regardless, we believe that the fact that such a short bout of therapy led to immediate treatment gains and generalisation effects for someone who was a relatively long time post-onset is encouraging and shows the potential benefits of the independent use of aphasia software and Action! in particular. Future research is needed to determine whether such treatment effects can be maintained for a longer period of time after termination of the therapy and which treatment regime might facilitate this. However, it is of note that it is plausible that JOG could have benefitted from a longer period of treatment, despite describing the therapy as "easy",

Participant DTR showed a different pattern across the treatment study, with her performance improving substantially across the three pre-treatment baseline assessments. As a result, she scored relatively close to ceiling at the final pre-therapy baseline, making any effect of treatment difficult to demonstrate. This was clear from looking at, for example, the improvement of lexical retrieval of treated items: While DTR did in fact significantly improve on this metric over the course of the treatment, the rate of improvement was not larger than that observed prior to the treatment. DTR's overall improvement during treatment was significantly different between the treated and untreated sets, both for overall accuracy as well as lexical retrieval separately.

The fact that DTR's performance on the control task remained stable across the baselines indicates that there was something inherent to the treatment-specific assessment that caused DTR's performance to rise sharply across the baseline assessments. One contributing factor to this pattern could be that the first assessment session was considerably longer in duration than the following sessions (we also tested her for two other Action! steps during the first assessment session so as to decide which one would be suitable for treatment). DTR indicated that she got tired towards the end of the session, which may have negatively affected her performance at this assessment. Descriptive statistics seemingly confirm this picture, with tense accuracy for the second half of assessment items (30.8%) markedly lower than that for the first half (51.4%) in this session. In addition, it is also possible that DTR became more familiar with the task over the course of the assessment sessions, as she learned what was expected of her in terms of the verb form that was required. This theory seems to be supported by the fact that it was tense (rather than lexical retrieval) that showed the greatest improvement after the first assessment session. It may be that the assessment process itself induced a learning effect for the use of tense (learning effects have previously been reported after brief exposure, see e.g. Byng, 1988). Clearly, however, for this hypothesis to be confirmed we would need to demonstrate generalisation beyond this specific task to another situation where DTR showed impaired tense marking prior to therapy. Without this demonstration, the pattern could be learning restricted to this task rather than an improvement in tense marking in general.

Given DTR's pattern of performance over the baseline period, the wisdom of including her in this study could be questioned. At the time of the first baseline assessment, there was plenty of room for DTR to show improvement, which is why we decided to treat her on past and future tense. Retrospectively, it may have been better to decide on a treatment step only at B2 or B3, although this would have involved many more lengthy assessments in order to decide on the most appropriate treatment step. While we were aware of the implications of DTR's upwards baseline performance going into the treatment, we felt that it remained important to include her in this study in order to gather her feedback on the usability of the app.

5.5.3 Action!'s usage, usability, and implications for future development and research

The current pilot study has provided us with useful data and feedback regarding Action! and its suitability for independent use by people with aphasia. The data collected by the app showed that both participants successfully used the app independently, with participant JOG far exceeding the prescribed treatment dose. The app data also showed that app users were able to personalise cue use, with one participant using cues extensively and the other not using cues at all (which may be due to her describing the treatment as "very simple" and therefore not needing any cues). Participants received minimal support during the treatment (being contacted by email or phone once to check on their progress) and relied on written instructions and an explanation and test run that they received prior to starting treatment. The fact that both participants were able to subsequently use the app independently is promising for Action!'s potential as an at-home therapy app.

The post-treatment interview further confirmed this impression, with both participants being very positive about Action!. Both indicated that they had enjoyed using the system, found it accessible and easy to use independently at home, and would recommend it to other people with aphasia. These results are particularly important as our participants showed rather different levels of impairment, indicating that the system could be useful for people with varying aphasia severities in terms of verb/sentence processing. It is important to note, however, that both participants had good comprehension, which may well be a prerequisite for successful independent use of Action and would need to be confirmed by further research. The accessibility aspect is of note here too, as both participants physically managed the app despite having reduced use of one hand, a common issue for people with aphasia. While both participants experienced some technical issues while using Action! (i.e. difficulty logging in and occasional app crashes), which will need to be resolved, it seems that Action! in its current form is a suitable tool for independent use by people with aphasia. Future research could help to further isolate the source of these initial technical hiccups and resolve them. It could also investigate the addition of different varieties of English (as currently the recordings are in Australian English). This should be a relatively easy feature to implement and would make the app more useful for English speakers outside of Australia.

Still, there are some issues that remain regarding Action!'s future development and its independent use by people with aphasia. Firstly, we cannot currently say which people (in terms of linguistic profile as well as aphasia severity) would benefit from using Action!. Even with just two participants in the current study, we noticed considerable variability in terms of treatment outcome. It is likely that the type of impairment, aphasia severity, non-linguistic cognition or other factors play a role here. Comprehension abilities (which were relatively intact for both of our participants) may also be of importance. It is unrealistic to expect that we could develop a precise prescription guide given the variability present in aphasia across many aspects of functioning and the difficulty determining which aspects are critical to guarantee success in any particular treatment (see e.g. Best & Nickels, 2000). Nevertheless, it would be beneficial to better understand which people are most likely to benefit from (which steps of) Action! treatment, in order to guide clinical decisions.

Secondly, we do not yet know the optimal treatment dose and intensity. While JOG's results show that Action! treatment can lead to improved outcomes even after a relatively short amount of therapy, we do not yet have sufficient information to formulate treatment guidelines. This issue is particularly relevant considering that JOG's improvements in tense marking were not maintained a month after the end of treatment. Treatment with longer duration (e.g. as provided by Links et al. (2010), which did lead to sustained improvements) may be a good starting point to address this issue.

Thirdly, we have currently only explored two of the Action! treatment steps. We have not looked at the efficacy of the remaining seven steps, which would be useful if we were to release the software for clinical use as it may help to guide clinical decisions. For all of these three issues, however, it is important to note that clinical decisions regarding software use and suitable treatment dosage/duration are ultimately made at the discretion of speech and language therapists, whose expertise and judgement will be the decisive factor, with Action! in its current form a potentially useful addition to their toolbox.

Fourthly, in its current form, Action! does not provide direct feedback to users based on their spoken responses. While they are provided with the correct answer, feedback is based on how they judge their own accuracy, which could potentially reinforce incorrect responses. We do not currently know whether this was the case for the participants in our study. Both participants judgements of their own accuracy were relatively high (73.7% of responses judged as accurate for JOG and 91.3% for DTR). For DTR, this would be in line with the feedback she provided at the post-treatment interview (describing the treatment as "very simple"), as well as with the fact that she did not use any cues for the duration of the treatment. However, users incorrectly rating their responses as correct could lead to app use in an unintended way (e.g. skipping cues which could lead to better performance).

Some degree of supervision by a clinician may therefore be desirable, at least initially to determine accuracy of judgements. This is feasible within Action! as all user interaction is stored, with audio recordings allowing clinicians to potentially monitor their clients' responses. We are currently conducting a research project to investigate how well JOG's self-ratings reflect his actual performance. In practice, this lack of direct feedback means that therapists may need to administer assessments to get a more objective measure of their client's performance, rather than rely on the Action! scores as a way of measuring improvement. Alternatively, therapists could use the audio recordings of users' responses that are automatically created by Action! or observe their clients during app use, which would give an indication of their level of performance. Automatic recognition of (in) correct responses by the software would be an ideal scenario, although this would require incorporating speech recognition that is capable of handling aphasic speech. Relatively recent research (e.g. Jacks et al., 2019) has highlighted the feasibility of this option, which may be facilitated by the fact that the Action! items require a specific target response.

Finally, an important issue (particular in the light of user-centred design; see e.g. Bannon (1986)) is that Action! does not yet have usability data from the second user group: Speech and language therapists. While we attempted to incorporate their needs as much as possible, through insights from previous research (see Swales et al., 2016), as well as our own survey (see Chapter 3), feedback from this user group is essential to ensure that our app meets the needs of clinicians as well as their views on the extent to which the app would meet the needs of the whole spectrum of people with aphasia.

We could also further explore most of these issues through conducting a larger-scale treatment study involving a case series of people with aphasia with varying patterns of impairment (although this would require substantial resources). Such research would also be an excellent opportunity to involve speech and language therapists and to collect more user feedback, which could be used to further improve the app in terms of user friendliness and accessibility.

However, even with these caveats in mind, Action! may still serve a clinical purpose in its current form. Importantly, the app is based on solid theoretical underpinnings and existing therapy programmes (e.g. Bastiaanse et al., 1997; De Aguiar et al., 2015; Furnas & Edmonds, 2014; Thompson et al., 2010; see Chapter 4 for more information) that have shown clinical benefits. While more research could help to determine Action!'s exact benefits, this evidence base provides reason for optimism with regards to the app's efficacy, particularly in combination with our results, which provide proof of principle that the app can lead to treatment effects. Furthermore, while in the current study Action! was used as a stand-alone treatment, it is well suited for use as a supplementary tool in addition to regular speech and language therapy. Homework exercises are an important element of speech and language therapy, and Action! provides a number of advantages over traditional pen-and-paper homework (e.g. use of cues, provision of the correct answer, user data available to clinician), making it a potentially useful addition to regular speech and language therapy in its current form. In fact, releasing the app in its current form as a beta-version for clinicians to use could simultaneously allow us to collect user feedback while also providing people with aphasia the opportunity to practice verb and sentence processing.

5.5.4 Conclusion

In the current pilot study, we trialled our newly developed Action! treatment app with two people with aphasia. While results differed across participants, there were some encouraging signs that Action! could be used effectively and independently for treatment of verb and sentence processing deficits. While more research is needed to better understand how Action! can be used as effectively as possible for a range of aphasia profiles, Action! in its current form could already be a useful addition to the clinician's toolbox. For now, these results are promising and add further evidence to the existing body of research showing that treatment software has considerable potential merits in aphasia therapy.