Delivering word retrieval therapies for people with aphasia in a virtual communication environment

Jane Marshall[1]*

Niamh Devane[1]

Lisa Edmonds[2]

Richard Talbot[1]

Stephanie Wilson[3]

Celia Woolf[1]

Nick Zwart[3]

[1] Division of Language and Communication Science, City, University of London, Northampton Square, London EC1V OHB, UK

[2]Aphasia Rehabilitation and Bilingualism Research Laboratory, Teachers College, Columbia University, New York, NY 10027, USA

[3]Centre for Human Computer Interaction Design, City, University of London, Northampton Square, London EC1V OHB, UK

*Corresponding author: J.Marshall@city.ac.uk

In Press: Aphasiology; DOI: 10.1080/02687038.2018.1488237

Abstract

Background: Delivering therapy remotely, via digital technology, can enhance provision for people with aphasia. EVA Park is a multi-user virtual island that can be used for such delivery. The first EVA Park study showed that daily language stimulation delivered via the platform improved functional communication and was positively received by users (Marshall et al, 2016; Amaya et al, 2018). This paper reports two single case studies, evaluating its capacity to deliver targeted language interventions. The first employed therapy for noun retrieval, using cued picture naming and modified Sematic Feature Analysis. The second employed modified Verb Network Strengthening Treatment (VNeST).

Aims: This study aimed to determine if treatment delivery was feasible in EVA Park, as assessed by participant compliance, treatment fidelity and participants' views. It explored the impact of the therapies on treated and untreated word production, connected speech and functional communication.

Methods & Procedures: Two participants with aphasia each received 20 sessions of individual therapy in EVA Park, delivered over 5 weeks. Feasibility was assessed by measuring compliance with the therapy regime, recording and checking the fidelity of 20% of treatment sessions, and using post therapy interviews to explore participant views. Treatment outcomes were evaluated via repeated measures single case designs, in which assessments were administered twice before therapy, immediately post therapy and five weeks later. Outcome measures included Object Picture Naming (study 1), Sentence Elicitation Pictures (study 2), Naming 84 items from the Object and Action Naming Battery (study 2), Narrative Production (Study 2), the Northwestern Assessment of Verb and Sentences: Argument Structure Production Test (Study 2) and Communication Activities of Daily Living – 2 (Study 1 & 2).

Outcomes & Results: Feasibility results were excellent. Both participants were fully compliant with the therapy regime. There was at least 90% fidelity with the treatment protocols and participant views were positive. Outcomes varied across the studies. The noun therapy significantly improved the naming of treated words, with good maintenance. Lexical gains were less evident on the Sentence Elicitation Pictures used in the VNeST study. Neither study demonstrated generalisation to untreated words, connected speech or functional communication.

Conclusions: Two treatment approaches, designed for face to face delivery, could be delivered remotely in EVA Park. Outcomes for the noun treatment were comparable to previous evaluations. Comparisons with previous research were more challenging for VNeST, owing to differences in methodology. Further evaluations of other treatment approaches are warranted.

Introduction

The amount of treatment provided to people with aphasia often falls short of what is recommended by research (Code & Heron, 2003; Code & Petheram, 2011; Katz, Hallowell, Code, Armstrong, Roberts, Pound, & Katz, 2000). Community services seem particularly constrained. For example, a survey of Australian providers found average regimes of 4 hours per week in inpatient rehabilitation hospitals, compared to just 1 hour per week in community settings (Verna, Davidson & Rose, 2009).

In the face of such concerns, some have called for more efficient and cost effective models of therapy delivery, including greater use of digital technology (Enderby, 2012). Computers can administer a range of speech, language and communication tasks, and so augment face-to-face therapy with self-directed practice (Palmer et al., 2012; Roper, Marshall & Wilson, 2016; Varley et al., 2016). They also allow for remote therapy delivery, with benefits for community services. For example, video conferencing technologies enable people with aphasia to undertake therapy in their own homes, while retaining live contact with their therapist, features which are highly valued by

service users (Law et al., 2010). There is encouraging evidence that such remote delivery is as effective as face-to-face therapy, at least for some forms of treatment (e.g. Woolf et al., 2016).

EVA Park offers a novel and engaging platform for remote therapy delivery. It is an online virtual environment (i.e. a simulated world that is accessed via the internet) that was created with people who have aphasia using a process of co-design (Wilson et al., 2015). It can be accessed from a person's home on a regular computer. Users are represented by personalised avatars, which allow all aspects of their appearance to be tailored. So hair, skin tone, body shape and clothing can all be selected. Communication occurs in real time mainly via speech, using head phones and a microphone. Users can also type using an Instant Messenger facility. EVA Park takes the form of a simulated island containing a number of settings, including houses, a café, a hair salon, a tropical bar and a disco. Thus communication practice can take place in locations that mimic the real world. There are further natural features, such as wild life, and fantastic elements, such as a planetarium, that aim to stimulate conversation and inject fun into therapy.

In the first application of EVA Park, 20 people with aphasia received a target 25 hours of language stimulation over 5 weeks. Participants met with a support worker in EVA Park for one hour a day, during which they undertook goal directed communication activities and conversation practice. Findings were very positive. Participants completed an average of just over 23 sessions, with no attrition. Outcome measures showed significant benefits for functional communication (Marshall et al., 2016) and interviews revealed very positive user views about the intervention (Amaya et al., 2018). Human computer interaction assessments showed that EVA Park could support a range of conversational exchanges, and was strongly associated with positive affect, e.g. marked by laughter, playfulness and joke making (Galliers et al., 2017).

These findings indicated that EVA Park was an effective tool for delivering the tested intervention. The lack of attrition compared very favourably with other studies of intensive (> 4 hours per week)

aphasia therapy (Brady, Kelly, Godwin, Enderby & Campbell, 2016), and was possibly related to the high levels of enjoyment reported in the participant interviews (Amaya et al, 2018). We hypothesised that the opportunities for situated practice, for example requesting a hair cut in a simulated salon, promoted the gains in functional communication and may bode well for the generalisation of therapy skills to everyday contexts.

The clinical potential of EVA Park would be enhanced by evidence that it can be used to deliver a range of interventions. Accordingly, our second application tested a number of individual therapy programmes targeting specific language skills. Using single case designs, five interventions were tested, all of which have research evidence showing effectiveness with 'conventional' delivery. Here we report findings from the two case studies that evaluated word retrieval therapies. The remaining studies evaluated therapies for sentence, discourse and narrative production. These will be reported elsewhere.

The first case study tested therapy for noun retrieval. The protocol combined cued picture naming tasks (Woolf et al., 2016) with an adapted version of Semantic Feature Analysis (SFA; Boyle, 2004). Both approaches have been shown to improve naming of the words directly targeted in therapy (e.g. Efstratiadou, Papathanasiou, Holland, Atchonti & Hilari, in press; Maddy, Capilouto & McComas, 2014; Nickels, 2002; Woolf et al., 2016). SFA, in particular, aims to boost activation in the semantic lexicon, with the hypothesis that this might promote generalisation of improvement beyond the treated set of words. Such generalisation has indeed been achieved in some studies (Lowell, Beeson & Holland, 1995), but not all (see review and arguments in Boyle, 2010). Generalisation of gains to connected speech is also variable, but has been demonstrated for both cued naming tasks (Conroy, Sage & Ralph, 2009) and SFA (Peach & Reuter, 2010). The tested treatment approaches have been deployed with people who have a range of aphasic profiles (e.g. see Maddy et al., 2014).

The second case study tested an adapted version of Verb Network Strengthening Treatment (VNeST; Edmonds, Nadeau & Kiran, 2009). This therapy also aims to improve lexical retrieval, but with a greater focus on verbs. It is premised on the theory that the meaning of a verb is partly dependent on the thematic roles with which it combines. Thus, the semantic representation of 'measure' includes information about its typical agents (carpenter, chef, surveyor) and patients (wood, flour, land). In line with this, the treatment aims to activate networks of verbs and their associated nouns, with the further hypothesis that related words, sharing members of the network, might also benefit. Evaluations of VNeST have supported this hypothesis. They have demonstrated improved production of treated verb/noun networks, with generalisation to untreated stimuli (Edmonds et al, 2009, Edmonds & Babb, 2011; Edmonds, Mammino, & Ojeda, 2014; Furnas & Edmonds, 2014). There were also encouraging improvements in discourse and in ratings for functional communication (e.g. see Edmonds et al, 2014). Although there was some individual variability in outcome, these studies showed that VNeST can be used successfully with people who have varying types and severities of aphasia.

This paper describes two case studies in which the above therapies were administered to an individual with aphasia in EVA Park. In both cases, treatment delivery capitalised on the opportunities for language stimulation inherent in EVA Park. Thus, in all sessions, formal therapy tasks were augmented by at least 10 minutes of situated practice in different EVA Park locations. This element aimed to promote the generalisation of target language skills into connected speech and functional communication. The case studies addressed the following research questions (RQs):

RQ1: Is delivery of the therapy protocol feasible in EVA Park, as assessed by participant compliance, fidelity checks on treatment videos and participant views?

RQ2: Does therapy improve the production of treated words in isolation (Case 1) or sentence contexts (Case 2)? Is there generalisation to untreated words?

RQ3: Do therapy benefits generalize to measures of connected speech and functional communication?

Case Studies

The case studies received ethical clearance from the Research Ethics Committee of the School of Health, City University of London (LCS/PR/Staff/16-17/04, July 2016). Participants gave informed written consent, using materials designed to be accessible to people with aphasia (Rose, Worrall, Hickson & Hoffmann, 2011).

Participant selection criteria were: chronic aphasia of over 4 months duration caused by a left hemisphere stroke; fluent premorbid use of English; not receiving speech and language therapy elsewhere for the duration of the study. To be eligible for case study 1 the participant had to score less than 50% correct on a 190 item noun naming assessment (Best et al., 2011). To be eligible for case study 2 the participant had to score less than 50% correct in producing verbs and nouns on a sentence elicitation task (Edmonds et al, 2009). Two participants, 'Blake' and 'Milton', were recruited via community stroke groups. Both were male (see other details in Table 1). Blake took part in study 1, and Milton in study 2.

Insert Table 1 about here

The studies employed a repeated measures design. Participants were assessed at four time points (T1 – T4), and each time point was separated by 5 weeks. Intervention was provided between T2 and T3, with no therapy received between the other time points. Thus the design offered a double baseline assessment, an assessment immediately post therapy and a five week follow up.

Testing, in both studies, was conducted face to face (i.e. not in EVA Park), by a non-treating therapist. All assessments were video recorded and transcription and scoring were conducted by

students or members of the research team who were not involved in data collection. For consistency, each test was scored by the same person at all time points. Videos were scored in random order and scorers were blinded to time point and treatment content.

Study 1 (Noun Retrieval Therapy) Method

Measures

RQ1

There were three indices of treatment feasibility. The first was compliance, i.e. the number of planned sessions attended by the participant. This was recorded in the therapists' notes. The second was treatment fidelity. Six therapy sessions (out of 20) were video recorded. Five recordings were selected for fidelity assessment. The selection was made randomly, i.e. without access to the content or order of the sessions. A fidelity checklist was compiled against the written treatment protocol (see checklists in appendix 1). Fidelity checking for both studies was conducted by a speech and language therapist who was not otherwise involved in the study. The third feasibility index was treatment acceptability. Blake was interviewed in the week after therapy completion to capture his views about the intervention. Interviews were conducted by a non-treating therapist, following the topic guide in appendix 2. With some questions, for example relating to the experience of therapy, he was given the option of providing a quality rating, ranging from 1 (not like) to 5 (like). The interview was recorded and transcribed and key themes identified.

RQ2:

The naming of 100 object picture names was assessed at each time point. Fifty of the object names were used as stimuli in the treatment tasks and 50 were untreated. Materials were drawn from a larger set of 190 items which have 95% name agreement when tested with healthy controls (Best et

al, 2011) and naming of these items has shown sensitivity to therapy induced change (Woolf et al, 2016). The treated and untreated words were matched for familiarity (treated group mean (sd): 522.22 (57.67); untreated group mean (sd): 524.84 (68.18); independent t test comparison, t = .201, p = .84) log word frequency (treated group mean (sd): 1.08 (.602); untreated group mean (sd): 1.08 (.605); independent t test comparison, t = .023, p = .98) and for baseline naming success (see results). During testing, Blake was shown each picture in turn and asked to name the item. No cues or corrections were provided. Production of the target name was scored correct, even if this followed an error. The order of presentation was changed on each testing occasion.

RQ3

We aimed to assess connected speech by collecting a narrative sample from Blake at each time point. The task involved retelling a personal story or event. However, Blake had very limited spontaneous speech and declined to carry this out.

The CADL-2 test (Holland et al, 1999) was administered at each time point to assess functional communication. This standardised assessment uses everyday situations, such as going to the doctor, to explore language use in context. Scoring reflects the ability to transmit information by whatever means, rather than formal language skills.

Therapy

Blake received 20, one hour sessions of therapy over 5 weeks. Three sessions per week were administered by a qualified speech and language therapist (ND). One session per week was administered by a final year student of speech and language therapy. All sessions were conducted remotely in EVA Park. Blake worked on a computer in his own home. The treating therapist worked on a computer in the University, or in their own home. Blake only met the treating therapist once face-to-face, when EVA Park was set up in his home. All other interactions took place in EVA Park.

About 10 nouns were treated per session, in a rotating order. The pictures were presented, one by one, on a screen in EVA Park. This was a blank screen installed for the purpose of therapy in an attractive location at the water's edge. The participant was asked two semantic verification questions about each item. For example, for lemon the questions were: 'Is it sweet?' and 'Can you squeeze it?' The participant was then asked to name the picture. If the naming attempt was correct, three repetitions were requested. If naming was not correct, the following series of cues was provided: a semantic cue ('it's sour'), a closure cue ('as sour as a ..'), a first phoneme cue (/l/), a first syllable cue (/lɛ/) and a repetition cue ('lemon'). Cues were discontinued when naming was achieved.

Modified Semantic Feature Analysis followed the cuing task. The screen in EVA Park showed the SFA elicitation chart. This showed the target picture in the centre, connected to six boxes, each of which had a category heading. Categories were: group, use, action, properties, location and association. Using the chart and appropriate questions, the therapist elicited at least one semantic feature under each category. For example, with the target 'lemon' the response for group might be 'fruit'. Features were verified by the therapist both orally and in writing, using EVA Park's Instant Messaging facility. If the participant was unable to name a feature, one was provided by the therapist for repetition. Figure 1 shows a screen shot of therapy in progress with Blake.

Insert Figure 1 about here

In both stages of therapy all questions, cues and feedback were delivered by the treating therapist. Like Blake, the therapist was represented by a personalized avatar in EVA Park and communicated remotely using headphones and a microphone. So, during sessions Blake saw the therapist's avatar in EVA Park and heard her voice over his headphones. Speech was not synthesized; i.e. Blake and

the therapist heard each other talking in real time. Avatars did not represent lip movements. Written feedback from the therapist, via the Instant Messaging facility, appeared in a box on Blake's computer screen.

Generalisation practice

The formal word finding tasks were followed by at least 10 minutes of informal language practice using different settings in EVA Park. This practice aimed to promote word production in context. Examples included: requesting items from fruit and flower stalls in the EVA Park town square, naming objects in the EVA Park houses, naming the plants in the EVA Park green house, and naming examples of EVA Park wild life. In order to support this practice, some EVA Park objects (such as the penguins) were adapted, so that they provided phonological cues for their name when clicked. These cues were hierarchical; i.e. one click elicited the first phoneme, while two elicited the whole word for repetition. Generalisation activities also involved talk about the real world. For example, Blake discussed foods that he likes and dislikes.

Study 1 Results

RQ1: Is delivery of the therapy protocol feasible in EVA Park, as assessed by participant compliance, fidelity checks on treatment videos and participant views?

Compliance with the therapy regime was excellent with all treatment sessions completed as planned. Fidelity findings were also strong. The five sessions that were checked for fidelity included 499 treatment components, of which 461 were compliant (92%). These components spanned both the formal word finding tasks and the generalisation practice. One of the checked sessions was delivered by the student. Fidelity of this session was excellent, with 99% compliance.

Blake's interview data (see examples in Table 2) revealed positive views about receiving therapy in EVA Park. He particularly valued the relationship with the therapist, which was given a maximum rating of 5. He also rated the generalisation practice at 5. The formal word finding tasks were given a lower rating of 3. This seemed due to the challenging nature of the tasks and his own difficulties with speech. Perhaps most encouragingly Blake felt that the therapy had made a difference to his communication. He gave the example of using a treated word to buy crabs from a fish stall (Blake lives by the sea).

Blake was very positive about the experience of using EVA Park. He rated his enjoyment at 5 and commented repeatedly that it was 'amusing'. He indicated places that he liked to visit (e.g. by drawing the EVA Park sail boat) and objects that he enjoyed. When asked about his avatar, Blake rated this at 5. As a joke, he had created an avatar with very long hair, in contrast to his real-life experience of baldness, and he laughed about this in the interview.

When asked about his feelings of control over the EVA Park technology Blake was generally very positive. For example, he described moving his avatar as 'easy'. He indicated that his computer occasionally crashed, but also showed that he was able to re-start it, by miming pressing buttons on the keyboard. More critically, he indicated that would value more interactive objects in EVA Park, i.e. objects that react when clicked. He also suggested that he would value more contact with other EVA Park users, e.g. when he visited the platform outside the treatment sessions.

Overall, the interview findings suggest that delivery of treatment via EVA Park was very acceptable to Blake. When asked if he would recommend this form of therapy for other people with aphasia he responded: 'yes yes yes' and gave a thumbs up gesture.

Insert Table 2 about here

RQ2 Does therapy improve the production of treated words? Is there generalisation to untreated words?

Table 3 shows the number of treated and untreated nouns that were named at each time point. Naming over the baseline period (T1 to T2) was stable. There was a significant improvement in the naming of treated words following therapy (T2 vs T3, McNemar χ^2 $p < .001$). This improvement was well maintained at T4 (T2 vs T4, McNemar χ^2 $p < .001$). Naming of the untreated words showed no change. Thus therapy improved the production of treated words, with no generalisation to untreated words.

Insert Table 3 about here

RQ3: Do therapy benefits generalise to measures of connected speech and functional communication?

Blake was unable to comply with the assessment of connected speech. Therefore only the CADL-2 data were available to address this question. These are reported in Table 4.

Although Blake's score increased over the therapy period (between T2 and T3), this was largely because of a decline at T2. Indeed his post therapy score is similar to the first baseline. This measure, therefore, showed little evidence of therapy induced change.

Insert Table 4 about here

In summary, study 1 showed that delivery of cued naming and modified SFA therapy was feasible in EVA Park. Therapy brought about improved naming of treated words, but no generalisation to untreated words. Functional communication, as assessed by CADL-2, was unchanged by therapy.

Study 2 (Modified VNeST) Method

Measures

RQ1

As in study 1, feasibility was assessed by recording the participant's compliance with the therapy regime and checking the fidelity of 5 randomly selected treatment sessions (see fidelity checklist in appendix 1). Treatment acceptability was explored in a post therapy interview conducted with Milton. The methods used for fidelity checking and the conduct of the interview were the same as in study 1.

RQ2

The production of treated and untreated verb networks was assessed by the Sentence Elicitation Pictures (SEP) employed by Edmonds et al (2009). Materials for the task consisted of 24 action pictures showing an agent acting upon a patient. The agents had clear roles so that specific rather than general vocabulary could be targeted, such as 'waiter' rather than 'man'. The verbs from 12 of the pictures were used as therapy stimuli. The other 12 verbs were untreated (see Edmonds et al, 2009 for details about matching). Each of the untreated verbs was semantically related to a verb in the treated set (e.g. chop/slice; push/pull).

During testing Milton was shown each picture in turn with the following instruction: 'I want you to make a sentence about this picture and include this (point to the agent), what she/he's doing (point to the action), and this (point to the patient)'. He was encouraged to use specific vocabulary. If he used a general term, like 'man' he was prompted to provide a more specific alternative. The order of presentation was changed at each time point, and related pictures were not presented consecutively. Milton's production was scored for the number of agents, verbs and patients produced. Thus the maximum score for each set of 12 pictures was 36, comprising 12 verbs and 24 nouns.

Naming of untreated verbs and nouns in isolation was also assessed with 84 items from the Object and Action Naming Battery (Druks and Masterson, 2000), extracted as documented in Edmonds, Obemermeyer and Kernan (2015). The sets of 42 verbs and 42 nouns were matched item-by-item (i.e., noun-verb pairs) for age-of-acquisition (paired t-test, $p > .05$). Edmonds et al. (2015) were not able to match the same pairs statistically on imageability. However, there was <1.0 rating difference across 36 of the 42 age-of-acquisition matched pairs. (See Edmonds et al., 2015 for the list of paired items). Milton was shown each picture in turn and asked to produce a single word to describe the depicted action or object.

RQ3

Two measures were used to assess connected speech:

i) A narrative sample was elicited at each time point. Milton was asked to describe something that was personally meaningful, for example by relaying an event that had occurred in the last week. Time was unconstrained. The sample was analysed using word and sentence level measures to evaluate lexical retrieval. Word level informativeness was evaluated by correct information units (CIUs) using the methods outlined by Nicholas & Brookshire (1993). Percent correct CIUs (%CIUs) reflects the percentage of words that were novel (i.e. not repeated) and relevant to the topic. Complete Utterances (CU) were evaluated to determine the percentage of utterances (%CU) that contained a complete subject, verb, object (SVO) sentence frame for which all the words were relevant to the topic and each other (e.g., Edmonds et al., 2014). The transcripts were broken into T-units, or a main clause with its subordinate clauses. Then, words, CIUs, and CUs were coded. A second researcher then coded the transcripts and marked any disagreements for designation of utterance breaks and coding for words, CIUs and CUs. Transcripts were presented in random order and all coding was conducted blind to time point. Initial reliability was 98.8% for utterance breaks, 99.2% for words, and 80% for CIUs and CUs. The initial reliability for CIUs and CUs is lower than

previous VNeST studies, but this is likely because previous studies used Nicholas and Brookshire (1993) stimuli, which are more constrained than the personal narratives provided in the current study. Subsequent to the initial reliability, the original 2 researchers reviewed and resolved disagreements through consensus, resulting in 100% agreement.

ii) The Northwestern Assessment of Verb and Sentences: Argument Structure Production Test (NAVS: ASPT, Thompson, 2011) was administered at each time point. As in previous evaluations of VNeST (Edmonds and Babb, 2011), a modified version of the test was administered. Milton was asked to describe 34 action pictures, illustrating one place (n = 8; e.g. 'the dog is barking'), two place (n = 16, e.g. 'the man is washing the clothes') or three place verbs (n = 10; e.g. 'the man is sending the letter to the woman'). In all pictures, arrows pointed to the argument nouns that should be included in the description. In a deviation from the original protocol, Milton was not provided with the target verb. Scoring credited each correct noun and verb produced by the participant. In addition, a total score of 1 was awarded for each sentence if all the required nouns and verbs were produced in the correct order.

The third measure used to address RQ3 was CADL-2 (Holland et al, 1999). This standardized assessment of functional communication in aphasia was administered at each time point.

Therapy

The therapy regime was the same as in case study 1 with 20 sessions administered over 5 weeks. As in study 1, one session per week was administered by a SLT student and the remaining sessions by a qualified therapist (RT). All sessions were administered remotely in EVA Park (see study 1 for further details on therapy administration).

Modified VNeST

The VNeST therapy worked on 12 treated verbs, in rotating order. Approximately 6 verbs were targeted in each session. Materials consisted of a sentence frame, which was projected on a screen in EVA Park, showing a verb, flanked by a box saying 'Who' and a box saying 'What'.

The therapist said the target verb and wrote it using the Instant Messenger facility. Three agents for the verb were elicited by indicating the 'Who' box and asking 'who' questions. Responses were verified verbally and in writing. Minimal cues involved further questions, such as 'who would boil things as part of their work?' Maximal cues involved offering potential choices for yes/no verification, such as 'would a chef boil things?' This is a minor adaptation of the published VNeST protocol, where maximal cues require selection of a plausible response in the presence of foils (e.g., Edmonds, 2014). Three patients were similarly elicited, e.g. by asking 'what would a chef boil?' and if necessary offering options for verification. Each completed triad was written down and Milton was asked to read them aloud, e.g. 'the chef boils soup'.

In early sessions, a sentence expansion frame was then displayed on the screen, showing 'Where', 'Why' and 'When' boxes. One triad was selected by Milton for expansion. Each expanding question was asked in relation to the chosen triad, for example: 'Where does the chef boil soup?' Milton's responses were written down and he was then required to read aloud an expanded sentence, containing all the elicited information: 'The chef boils soup in the restaurant, for his customers at dinner time'. Expansion continued in the later sessions, but the frame was rarely needed.

Previous studies of VNeST included a sentence judgement task, in which participants were asked to detect anomalous verb/noun combinations. In consultation with the therapy originator (LE) this was omitted from the current study, to allow time for the generalisation practice described below. The final formal task, therefore, involved independently producing each treated verb upon request. If Milton was unable, he was encouraged to think of associated agents and patients; and if that failed he was cued with the written word.

Generalisation practice

As in study 1, the formal language tasks in each session were followed by at least 10 minutes of informal language practice using different settings in EVA Park. This aimed to generalise the treated skills into connected speech and functional communication. Examples for case study 2 included: discussing actions that could take place in different EVA Park locations, and listing the possible activities of EVA Park animals. To support the practice, 12 objects in EVA Park were animated to enact the treatment verbs. For example, the kettle in one of the EVA Park houses boiled when clicked and the brush could be used by avatars to clean the floor. Real world conversation topics included activities undertaken over the weekend and (disreputable) actions that could occur at a party.

Study 2 Results

RQ1: Is delivery of the therapy protocol feasible in EVA Park, as assessed by participant compliance, fidelity checks on treatment videos and participant views?

Compliance was excellent, with no missed sessions. Fidelity findings were also strong. The 5 checked sessions included 344 treatment components, of which 312 were compliant (93%). One student led session was included in the sample. This was 90% compliant.

Milton expressed positive views about receiving therapy in EVA Park (see Table 2 for example quotes). He was content to work with the therapist remotely and, in a deliberate violation of our scale, rated the interactions with the therapist as 'ten out of ten'. In terms of therapy content, he indicated that the focus of therapy was very appropriate for his needs. He also seemed to value the meta-linguistic nature of the tasks, and the way that therapy invited him to reflect on the structure of language. Less positively, he said that he found the therapy monotonous and would like to work on a larger range of verbs. When asked about the generalisation practice he indicated that he would

like to do more of this. He commented on the experience of first practising verbs in the formal tasks, and then attempting to access the same vocabulary in relevant EVA Park settings. His conclusion that it was 'all joined up' suggested that this gave a particular cohesion to the therapy.

Milton rated his enjoyment of EVA Park highly (five out of five). He valued the novelty of the platform: 'a nice feeling to do it slightly different' and the fact that it was exciting and funny. He cited locations that he liked to visit and EVA Park animals that he enjoyed. He was happy being represented by an avatar and found it easy to navigate the island. When asked about technological aspects the only problem referred to was an occasional freezing of the number lock on the key pad.

In terms of treatment impact, Milton felt that therapy improved his access to vocabulary and increased his success rate in word-finding from 50 to 70%. He also suggested that he was applying some of the acquired metalinguistic skills in his everyday conversations.

These interview findings suggest that delivery of treatment via EVA Park was very acceptable to Milton. He, like Blake, said he would recommend this form of therapy for other people with aphasia.

RQ2 Does therapy improve the production of treated words? Is there generalisation to untreated words?

Table 5 shows the number of verbs and nouns named from the Sentence Elicitation Pictures at each time point, with the break down for treated and untreated verbs. Considering verbs first, total production was stable across the baseline period. There was a small increase following therapy (T3), which was maintained at T4. The change was not significant (T2 vs T3, McNemar χ^2 p = 0.18; T2 vs T4, McNemar χ^2 p = 0.18). Both treated and untreated verbs improved, but very marginally. Total noun production increased between T1 and T2 (McNemar χ^2 p = 0.007) and then stabilised across T3 and T4. These results indicate that therapy did not significantly improve the production of treated or untreated words.

Insert Table 5 about here

Table 6 reports scores on the Object and Action naming battery. Taking the total scores, there was a marginal, but non-significant improvement over the baseline period (T1 vs T2, McNemar $\chi^2 = 2.1$, $p>.05$). At T3, there was a further gain that now reached significance (T2 vs T3, McNemar $\chi^2 = 4.57$, $p<.05$). This gain was distributed equally over the nouns and verbs. The total score at T4 was no longer significantly better than the second baseline (T2 vs T4, McNemar $\chi^2 = 3.6$, $p>.05$). Thus, there was a significant increase in naming following therapy that was not maintained at follow up.

Insert Table 6 about here

RQ3: Do therapy benefits generalise to measures of connected speech and functional communication?

Insert Table 7 about here

Data for the narrative transcripts are presented in Table 7. A positive response to RQ3 would be indicated by an increase at T3 in the %CIU and %CU. This is not shown. Rather scores declined between T2 and T3, albeit rallying at T4. The highly variable sample size should be noted. For example Milton produced 1017 words at T1 compared to 276 words at T2. This makes the scores difficult to interpret. It was also not possible to compare Milton's scores to the normative data in Nicholas and Brookshire (1993) as different elicitation procedures were used.

Scores on the NAVS Argument Production Test (Thompson, 2011) are shown in Table 8. These scores do not offer evidence of improvement. In part, this was due to ceiling effects, as verb production was surprisingly high even at baseline. There was a margin for change with the total sentence scores, which were awarded if the verb and all required nouns were produced, and in the correct order. Yet there was not a significant gain on this score.

Insert Table 8 about here

Results on the CADL-2 assessment are outlined in Table 4. Milton improved at each assessment point, regardless of therapy. Indeed the greatest change occurred over the two baseline assessments. His scores do not provide strong evidence that therapy improved functional communication.

In summary, case study 2 showed that delivery of modified VNeST was feasible in EVA Park. Treatment effectiveness, as assessed by our measures, was not shown.

Discussion

This paper reported two single case studies of language therapy delivered in a virtual environment called EVA Park. It aimed to explore whether the treatment approaches, which were developed for face to face delivery, can be adapted for this environment, and to evaluate the outcomes from those approaches. Here we discuss the findings against each of our research questions, before considering future implications.

RQ1 asked whether delivery of the treatment protocols was feasible in EVA Park. Findings were very positive. Both participants were fully compliant with the therapy regime with no missed sessions. Fidelity scores were also good, showing that virtual delivery did not induce drift from the treatment protocol. This was the case regardless of whether therapy was conducted by a qualified therapist or student, indicating that EVA Park might also be employed for delegated treatment delivery. Participants' views about the experience of receiving therapy in EVA Park were also very positive. In line with our previous study (Amaya et al, 2018) there was a strong theme of enjoyment and reports of a warm rapport with the therapist, which was not harmed by the remote and virtual nature of the interaction. The opportunity for generalisation practice in the simulated environments of EVA Park was also valued. There were criticisms of the treatment content and the technology. Some of these

related to the therapy protocol, so would presumably also have arisen from conventional delivery. Problems relating to technology could be resolved, and did not sabotage the participants' overall experience.

Results for RQ2 differed across the two case studies. In case study 1 therapy produced a highly significant and well-maintained improvement in the naming of treated words. However, this gain did not extend to untreated words, where naming was stable across the four assessment points.

These results compare favourably with previous studies that have employed face to face delivery of noun retrieval therapies. For example, a recent systematic review of the literature relating to Semantic Feature Analysis (Efstratiadou et al., in press) identified 19 studies, reporting a total of 47 participants. Across all studies, 37 participants improved in the naming of treated items, and 28 maintained their gain. In contrast only 14 participants achieved generalisation to untreated items. The treatment dose across the reviewed studies was similar to that employed here (mean: 16.9 hours, range 25 - 9 hours). One previous study of remote naming therapy, using video conferencing technology, also reported similar gains to those achieved here, in that treated words improved, with minimal generalisation to untreated items (Woolf et al., 2016).

Interpretation of results for case study 2 is more complex, partly because none of the tests evaluated treated words in isolation. Rather all required a degree of generalisation, for example to sentence contexts or untreated words. Further, there is not an expectation that all participants receiving VNeST will improve on all measures, given the potential effect of pre-treatment impairment profiles and the different degree of constraints across tasks (see Edmonds et al., 2015 for more details). The Sentence Elicitation Pictures and naming of items from the Object and Action Naming Battery are fairly constrained tasks, as there is little variability in the responses that can be offered. There were no significant changes on the Sentence Elicitation Pictures following therapy for picture stimuli that contained trained or untrained verbs. Total naming on the 84 items from the

Object and Action Naming Battery was fairly high at pre-treatment and did improve at T3. However, the change was similar in magnitude to that which occurred over the baseline period, and was not highly significant, raising concerns about a type 1 error. It was also not maintained at T4.

A recent review summarises the results achieved with 22 individuals across 6 studies of VNeST (Edmonds, 2016). This found that 86% of participants improved in the naming of nouns, and 58% in the naming of verbs. So, there is good evidence that VNeST improves lexical retrieval when delivered face to face. Accounting for the different results achieved here is difficult. Previous studies showing change on the Sentence Elicitation Pictures employed repeated probing, to establish stable baselines and evaluate the effects of the treatment phase. Such probing was not possible in the current study, given that testing was administered face to face and therapy remotely. It is possible that practice effects from the probing may have inflated previous treatment outcomes (e.g, Boyle, 2000), though control tasks probed at the same time points did not change with treatment. There was some modification to the VNeST protocol in this study, which may have influenced the result. Dosage also differed between this study and previous VNeST studies, a point on which we will elaborate below.

RQ3 asked whether treatment would improve connected speech and functional communication. Connected speech was not measured in case study 1, as Blake could not comply with the narrative task. However, observation and his continued inability to attempt a narrative suggested that connected speech remained highly problematic. Milton, in case study 2, was assessed on both a narrative and a sentence production task, with neither showing evidence of change. Evaluation of the former was hampered by the variable word count produced at each assessment point. Previous evaluations of VNeST showed discourse changes on standard elicitation procedures, such as complex picture description (Edmonds et al, 2009). Adopting these procedures in the current study may have reduced the variability and been more sensitive to change. Milton's lack of change on the sentence production measure also contrasts with previous investigations of VNeST (Edmonds, 2016). It is of interest that Milton had difficulties with thematic role assignment in sentence production (e.g.,

reversing the subject and object), which particularly affected his NAVS scores. So, he was able to produce all the lexical items in some sentences but had difficulty with the order of animate thematic roles, for example where a sentence involved a cat and a dog. Although VNeST promotes the use of SVO sentences, animate patients/direct objects are rarely produced due to the nature of the treated verbs. Thus Milton was rarely challenged to assign two animate roles to an SVO structure during his therapy. The sentence judgment step, normally conducted with VNeST, requires participants to judge the correctness of presented sentences, including sentences with thematic role reversal. This might have addressed Milton's problem. However, we deliberately removed this step to allow more time for generalisation tasks in EVA Park. In retrospect, including this step might have been beneficial. Findings with respect to functional communication as assessed by the CADL-2, did not offer evidence of gain for either participant, mainly because baseline measures were unstable.

The generalisation of treatment benefits to discourse and real world communication is a key goal for aphasia intervention (Brady et al, 2016). However, this is acknowledged to be difficult to measure and difficult to achieve (see arguments in Milman, 2016; Webster, Whitworth & Morris, 2015). The inclusion of generalisation practice, in which the target language skills could be employed in connected speech and different EVA Park contexts, aimed to promote generalisation in the current study. It is therefore disappointing that gains were not seen. Giving more time to this component of therapy might have achieved better results. The sensitivity of the measures may have been a further problem, especially as both participants felt that therapy *had* changed their everyday communication.

We return now for a deeper examination of the VNeST dose and delivery in EVA Park. A recent review of VNeST studies indicates that most participants received at least 3 hours of therapy per week for ten weeks (Edmonds, 2016). This dose of at least 30 hours contrasts with our regime of 20 hours delivered over 5 weeks. The time spent on core VNeST tasks was further reduced by our addition of EVA Park generalisation activities. We should also note that the EVA Park therapy trained

12 verbs, in contrast to previous VNeST studies which trained 10. A higher therapy dose, combined with a smaller treatment set allows for more time spent per verb. The clinician can therefore target diverse responses for each verb, and responses that are of high personal relevance. This aims to increase participant engagement and broaden the range of semantic networks activated, both of which have been hypothesized to contribute to generalisation (Edmonds, 2014, 2016). It is possible that training more verbs over fewer hours in EVA Park did not allow such deep semantic engagement to occur. Milton's feedback indicated that he found aspects of the therapy monotonous, and that he wanted to work on more verbs. Paradoxically, more therapy time might have promoted more diverse and meaningful responses, and hence reduced the monotony. However, this is only supposition, as there have been improvements reported with lower dosages of VNeST (e.g., Edmonds et al., 2009). It is also unclear whether the lack of face to face interaction was a factor in our results. Previous VNeST studies have all involved face to face delivery, although one provided this through teletherapy (Furnas & Edmonds, 2014). It is unsurprising that such questions remain after a pilot study bridging different treatment protocols. Further investigations of VNeST in EVA Park would be merited, particularly given the good feasibility results and favorable feedback from Milton.

The overarching aim of this study was to investigate whether EVA Park could be used for the remote delivery of treatment protocols, targeting specific aspects of language. The answer to this question was 'yes'. Findings for RQ1 showed that participants could comply with therapy and found it highly acceptable. As in our previous study, both reported instances of fun and enjoyment, suggesting that the motivating aspects of EVA Park were felt even in the context of formal therapy. The data on fidelity also showed that the treatments could be delivered as intended, despite the virtual context. The outcome data for case 1 showed that noun retrieval therapy delivered in EVA Park brought about gains that were comparable to those reported from face to face delivery. Findings from case 2 were more disappointing. However, it would be premature to attribute these to the delivery

format, given that outcomes across all VNeST studies have varied, and given that there were important differences between our design and the designs of previous studies.

The results of this study indicate that EVA Park might be employed by practising clinicians to deliver individual programmes of language therapy. This may be particularly relevant for clients in remote areas, or where travel to outpatient centres is not possible. The fact that 20% of the treatment sessions were delivered by a SLT student suggests that EVA Park treatment delivery can also be delegated, at least in part, to non-qualified staff. Further evaluations of SFA and VNeST would be merited in EVA Park. Feasibility for other treatment approaches also needs to be tested and will be reported in future papers.

Acknowledgements

This study was funded by The Tavistock Trust for Aphasia. EVA Park was originally developed with funding from the Stroke Association, grant number TSA 2011/10. We thank 'Blake' and 'Milton', our participants with aphasia, who generously gave their time to this research. We thank Wendy Best who made her test materials available to us. We thank the following students and therapists who contributed to this study: Beth Ann Ingrassia, Kimberley Maguire, Suzanna Mackenzie-Smith, Chloe Paddock, Ruth Ryder, Lotta Urban and Angela Watkins.

References

Amaya, A., Woolf, C., Devane, N., Galliers, J., Talbot, R., Wilson, S., Marshall, J. (2018) Receiving Aphasia Intervention in a Virtual Environment: The Participants' Perspective. Aphasiology, DOI 10.1080/02687038.2018.1431831

Berndt, R., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000) Quantitative production analysis: A training manual for the analysis of aphasic sentence production. Hove: Psychology Press

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 and Rehabilitation, 33, 229–242.

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.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, *13*(3), 236-249.

Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. (2016) Speech and language therapy for aphasia following stroke. Cochrane Database of Systematic Reviews, Issue 6. Art. No.: CD000425. DOI: 10.1002/14651858.CD000425.pub4

Code, C., & Heron, C. (2003). Services for aphasia, other acquired adult neurogenic communication and swallowing disorders in the United Kingdom, 2000. Disability and Rehabilitation, 25, 1231–1237.

Code, C., & Petheram, B. (2011). Delivering for aphasia. International Journal of Speech-Language Pathology, 13, 3–10.

Conroy, P., Sage, K. and Ralph, M. L. (2009), Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech?. International Journal of Language & Communication Disorders, 44: 1036.

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

Edmonds, L. (2016) A review of Verb Network Strengthening Treatment: Theory, methods, results and clinical implications. Topics in Language Disorders, 36, 2, 123 – 135.

Edmonds,L. Mammino, K. & Ojeda, J. (2014) Effect of Verb Network Strengthening Treatment (VNeST) in Persons With Aphasia: Extension and Replication of Previous Findings American Journal of Speech-Language Pathology, 23, S312–S329

Edmonds, L. A., & Babb, M. (2011). Effect of Verb Network Strengthening Treatment in moderate-to-severe aphasia. American Journal of Speech-Language Pathology, 20, 131–145.

Edmonds, L. A., Nadeau, S., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. Aphasiology, 20, 644–675.

Edmonds, L. A., Obermeyer, J., & Kernan, B. (2015). Investigation of pretreatment sentence production impairments in individuals with aphasia: Towards understanding the linguistic variables that impact generalisation in Verb Network Strengthening Treatment. Aphasiology, 29(11), 1312–1344.

Efstratiadou E.A., Papathanasiou I., Holland R., Archonti, A., Hilari K. (in press) A systematic review of Semantic Feature Analysis therapy studies for aphasia. Journal of Speech, Language and Hearing Research.

Enderby, P. (2012) How much therapy is enough? The impossible question! International Journal of Speech-Language Pathology, 2012; 14(5): 432–437.

Furnas, D. W. & Edmonds, L. A. (2014) The effect of computerised Verb Network Strengthening Treatment on lexical retrieval in aphasia, Aphasiology, 28:4, 401-420, DOI: 10.1080/02687038.2013.869304

Galliers, J., Wilson, S., Marshall, J., Talbot, R., Devane, N., Booth, T., Woolf, C. & Greenwood, H. (2017) Experiencing EVA Park, A Multi-User Virtual World For People With Aphasia. Transactions in Accessible Computing. TACCESS, 10 (4).

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

Katz, R. C., Hallowell, B., Code, C., Armstrong, E., Roberts, P., Pound, C. and Katz, L. (2000), A Multinational Comparison Of Aphasia Management Practices. International Journal of Language & Communication Disorders, 35: 303–314.

Law, J., Huby, G., Irving, A-M., Pringle, A-M., Conochie, D., Haworth, C. & Burston, A (2010) Reconciling the perspective of practitioner and service user: findings from The Aphasia in Scotland study. International Journal of Language & Communication Disorders, 45, 5, 551-560.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The Communicative Effectiveness Index: Development and psychometric evaluation of a functional communication measure for adult aphasia. Journal of Speech and Hearing Disorders, 54, 113–124.

Lowell, S., Beeson, P. M., & Holland, A. L. (1995). The efficacy of a semantic cueing procedure on naming performance of adults with aphasia. American Journal of Speech-Language Pathology, 4, 109–114

Maddy, K. M., Capilouto, G. J., & McComas, K. L. (2014). The effectiveness of semantic feature analysis: An evidence-based systematic review. *Annals of physical and rehabilitation medicine*, *57*(4), 254-267.

Marshall, J., Booth, T., Devane, N., Galliers, J. R., Greenwood, H., Hilari, K., Talbot, R., Wilson, S. & Woolf, C. (2016). Evaluating the Benefits of Aphasia Intervention Delivered in Virtual Reality: Results of a Quasi-Randomised Study. PLoS One, 11(8), e01603. doi: 10.1371/journal.pone.0160381

Milman, L. (2016) An integrated approach for treating discourse in aphasia bridging the gap between language impairment and functional communication. Topics in Language Disorders, 36, 1, pp. 80–96

Nicholas, L. & Brookshire, R. (1993) A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults With Aphasia. Journal of Speech, Language, and Hearing Research, 36, 338-350. doi:10.1044/jshr.3602.338.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising and reviewing. Aphasiology, 16(10–11), 935–979.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., Dimairo, M., Dixon, S., Mortley, J., Hilton, R,, Delaney, A. & Hughes, H. (2012) Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial, Stroke, 43, 7, 1904-11.

Peach, R. K., & Reuter, K. A. (2010). A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology*, *24*(9), 971-990.

Roper, A., Marshall, J. & Wilson, S. (2016). Benefits and Limitations of Computer Gesture Therapy for the Rehabilitation of Severe Aphasia. Frontiers in Human Neuroscience, 10, 595.. doi: 10.3389/fnhum.2016.00595

Rose, T., Worrall, L., Hickson, L. & Hoffmann, T. (2011) Aphasia friendly written health information: Content and design characteristics, International Journal of Speech-Language Pathology, 13:4, 335-347.

Thompson, C. K. (2011). Northwestern Assessment of Verbs and Sentences. Evanston, IL.

Varley, R., Cowell, P. E., Dyson, L., Inglis, L., Roper, A., & Whiteside, S. P. (2016). Self-Administered Computer Therapy for Apraxia of Speech: Two-Period Randomized Control Trial with Crossover. Stroke, 47 (3), 822-828.

Verna, A., Davidson, B. & Rose, T. (2009) Speech-language pathology services for people with aphasia: A survey of current practice in Australia. International Journal of Speech-Language Pathology, 11 , 3, 191 – 205.

Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop “fishing”? A review of generalisation following aphasia intervention. Aphasiology, 29(11), 1240–1264.

Wilson, S., Roper, A., Marshall, J., Galliers, J. R., Devane, N., Booth, T. & Woolf, C. (2015). Codesign for People with Aphasia Through Tangible Design Languages. CoDesign, 11(1), pp. 21-34

Woolf, C., Caute, A., Haigh, Z., Galliers, J. R., Wilson, S., Kessie, A., Hirani, S. P., 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), pp. 359-373.

Appendix 1: Fidelity Checklists

Noun Retrieval Therapy	
1	Item shown on board and name invited
2	If correct name is produced this is repeated 3 times
3	If the incorrect name is produced the cuing hierarchy is followed and the word is repeated 3 times when named
4	Semantic Feature Analysis chart shown
5	Group feature(s) elicited
6	Use feature(s) elicited
7	Action feature(s) elicited
8	Properties elicited
9	Location(s) elicited
10	Association(s) elicited
11	Features confirmed in writing (using Instant Messaging) by clinician or provided for repetition
Generalisation practice	10 minutes of conversation at the end/during the session in different areas of EVA Park
Generalisation practice	Treated words targeted/elicited in conversation

VNeST	
1	The therapist says the verb and writes it using Instant Messaging
2	3 possible agents are elicited
3	3 possible patients are elicited
4	The therapist uses minimal cues (eg 'who might do this for their job') or maximal cues (choices for selection) to support the participant
5	The therapist uses Instant Messaging to write the agents and patients round the verb
6	The therapist and participant read each completed triad aloud

7	One triad is chosen for expansion
8	Where is elicited
9	When is elicited
10	Why is elicited
11	The expanded sentence is read aloud (original triad plus where, when, why additions)
12	The participant is asked to name the target verb; if unable he is cued by being told to think of the triads, or with the written word
13	specific terms (e.g. ‘chef’ rather than ‘man’; ‘sausages’ rather than ‘dinner’) are encouraged/elicited
14	Different meanings of the verbs are employed in the triads
Generalisation practice	The participant is taken to different part(s) of EVA Park for conversation practice
Generalisation practice	Treated verbs are targeted/elicited in the conversation practice

Appendix 2: Topic Guide for post therapy interviews

The experience of EVA Park

How much did you enjoy being in EVA Park? (Rating scale offered)

Was there anything you particularly liked?

Was there anything you particularly disliked?

Was there anything you found difficult? If so what?

How do you find being in EVA Park on your own? (Rating scale offered)

Is there anything you didn't like or found boring about EVA Park when you are on your own?

You can interact with some things in EVA Park (e.g. sit on the chairs, lie in the bath, make the donkey bray). What do you think of this?

Did you interact with other people with aphasia in EVA Park? If so how did you find that?

The relationship with the therapist

How do you find being in EVA Park with [name of therapist]? (Rating scale offered) e.g. rapport, support, interaction

The content and impact of therapy

What do you think of therapy in EVA Park? (Rating scale offered), e.g. tasks, activities, generalisation practice?

Do you think EVA Park therapy made a difference to your communication? If so in what way?

Have you used things you practiced in EVA in your daily life? If so, can you give me an example?

The technology

How do you find moving your avatar? (Rating scale offered)

How do you find using the keypad? (Rating scale offered)

How do you find using the mouse (or trackpad)? (Rating scale offered)

What do you think of your avatar? (Rating scale offered)

Tables

Table 1: Participant Details

	Age at Recruitment (years)	Stroke Information	Time post Stroke (months)	Pre-stroke Occupation	Aphasia
Blake (Case study 1)	60	Left Hemisphere, Right Hemiplegia	60	Chemist	Moderate/severe non-fluent
Milton (Case study 2)	54	Left Hemisphere, no hemiplegia	36	Managing Director	Moderate fluent

Table 2: Themes identified in the post therapy interviews with illustrative quotes

Theme	Quote
Relationship with therapist	'How did you find being in EVA Park with (name of therapist)?' 'five (out of five) … Amusing' (Blake, Case study 1) 'How did you find being in EVA Park with (name of therapist)?' 'Oh I'm going to have to go ten, er, ten out of ten' (Milton, Case study 2)
Therapy content	'What did you think of the therapy in EVA Park?' 'Three (out of 5) .. you know three but (gestures speaking) speaking oh, don't like' (Blake, Case study 1) 'It was great for me, it was exactly what I want to have' (Milton, Case study 2) 'I got my, got my tool like. That's a verb. Ah that's what .. verb, noun adjective, all that sort of stuff. Ah that's what it is. It's all sorted' (Milton, Case study 2) 'Erm monotolus .. monotolus …[..] .. I think, oh why can't we do a different verb' (Milton, Case study 2)
Generalisation practice	'yeah, yeah practice (indicates rating of five)' (Blake, Case study 1) 'I realise that some of my verbs and it's er, and it's, you know, "pull" and "boil" and all that sort of stuff. I realised that actually when I got .. when I talk about "boil" and all the different, er, things I do for boil .. then we go go er [..] … in the house for example, er, and the kitchen [..] I can work out [..] gosh er, it's um it's er joined up' (Milton, Case study 2)

Impact on communication	‘so thing you practise in EVA, have you used in your .. in the real world? ‘yeah’ ‘What can you give me an example?’ ‘erm crabs’ (Blake, Case study 1) ‘when I have a conversation I know what chunk is, is, is what it’s all … how it’s working’ (Milton, Case study 2) ‘before .. it was like 50% but now it’s sort of 70%’ (Milton, Case study 2) ‘my vocabulary is, is a little bit better’ (Milton, Case study 2)
Fun and Enjoyment	‘musing .. amusing’ (Blake, Case study 1) ‘’It’s exciting and it’s funny’ (Milton, Case study 2)
EVA Park features and technology	‘What did you particularly like?’ [..] ‘yeah finding things … (thumbs up gesture) wow’ (Blake, Case study 1) ‘there was a lot of things that we can do ..[..] .. you went to this, this, the underground er, the sea and all that sort of stuff’ (Milton, Case study 2)

Table 3: Case Study 1 (Blake) Naming of treated and untreated nouns (number correct) at the four time points

	Treated (n=50)	Untreated (n=50)	Total (n=100)
Time 1	28	27	55
Time 2	25	27	52
Time 3	44	25	69
Time 4	41	27	68

Table 4: Results on the CADL-2 Assessment

	T1	T2	T3	T4
Blake (Case 1)	85	77	88	90
Milton (Case 2)	71	80	86	92

Table 5: Case Study 2 (Milton) Verb and noun production from the Sentence Elicitation Pictures at the four time points

	Treated Verbs		Untreated Verbs		Total	
	Verbs (n=12)	Nouns (n=24)	Verbs (n=12)	Nouns (n=24)	Verbs (n=24)	Nouns (n=48)
Time 1	4	12	7	8	11	20
Time 2	6	16	5	15	11	31
Time 3	9	16	7	16	16	32
Time 4	8	16	8	14	16	30

Table 6: Case Study 2 (Milton) Scores on the Object and Action Naming Battery

	Verbs (n=42)	Nouns (n=42)	Total (n=84)
T1	32	32	64
T2	35	36	71
T3	39	40	79
T4	37	40	77

Table 7: Case Study 2 (Milton) Data from the narrative task

	Sec	W	%CIU	%REL	%SVO	%CU	CIU/MIN	%Pauses
T1	520	1017	26.75	31.18	87.10	29.03	26.75	7.12
T2	154	276	38.77	48.39	70.97	41.94	38.77	12.34
T3	535	1367	28.46	38.98	91.53	38.98	28.46	1.50
T4	315	796	43.84	53.16	84.81	51.90	43.84	2.86
T1+T2 (mean)	337	646.5	32.76	39.78	79.03	35.48	32.76	9.73
T3+T4 (mean)	425	1081.5	36.15	46.07	88.17	45.44	36.15	2.18

Key:

Sec=seconds; W=Words according to Nicholas & Brookshire (1993); %CIU = Correct Information Units / Total Words x 100, Word-level measure of informativeness; %REL = number utterances relevant to the topic / Total utterances x 100, Utterance level measure of the informativeness without consideration of syntactic completion; %SVO = number of utterances which contain an SVO / Total utterances x 100, Measure of the ability to provide an SVO regardless of meaning; %CU = complete utterances (that have an intact SVO, and all the words in that SVO are relevant the subject) / Total utterances x 100; CIU/MIN = CIUs per minute; % Pauses = Total pause time, where pauses of 2 seconds or more were noted / Total time x 100.

Table 8: Case Study 2 (Milton) Verb, noun, and total sentence scores on the Northwestern Assessment of Verb and Sentences: Argument Structure Production Test

	Verb (n=34)	Subject Noun (n=34)	Direct Object Noun (n =26)	Indirect Object Noun (n=10)	Total Sentence Score (n=34)
T1	34	28	14	8	19
T2	34	29	11	9	16
T3	34	28	16	9	20
T4	33	29	16	7	21

Figure 1

Modified Semantic Feature Analysis delivered in EVA Park: Blake (left) and treating therapist (right)