


RESEARCH REPORT

Treatment fidelity of technology-enhanced reading therapy (CommuniCATE) for people with aphasia

Katharine Bacon¹  | Jane Marshall² | Anna Caute³ | Katie Monnelly² | Madeline Cruice² | Corinne Moutou¹ | Celia Woolf²

¹ City University of London, London, UK

² Department of Language and Communication, City University, London, UK

³ School of Health and Social Care, University of Essex, Colchester, UK

Correspondence

Katharine Bacon, City University of London, London, UK.

Email: katharine.bacon@city.ac.uk

Abstract

Background: Treatment fidelity (TF), that is, the degree to which the treatment delivery has adhered to protocol, is an important aspect of establishing treatment validity and reliability. Research has shown that establishing TF is only done in a small percentage of aphasia treatment studies.

Aims: This project supports the work of the CommuniCATE study, which explored the benefits of technology-enhanced aphasia therapy on participants' reading, writing, speech and conversation skills. It examines the TF of the Reading strand of the CommuniCATE project by assessing whether the therapy adhered to the protocol. The following research questions were asked: Does treatment delivery adhere to treatment protocol? Does the degree of TF vary according to the person delivering the therapy (i.e. student therapist or qualified therapist)? Does the degree of TF vary over time (early treatment sessions compared with later treatment sessions)? Was the checklist tool reliable?

Methods & Procedures: This study assessed the fidelity of 38 retrospective video recordings of therapy. It used a checklist measure of criteria to which the delivery of the sessions should adhere, and against which the sessions were rated. Participants were the people with aphasia receiving therapy, the students and qualified speech and language therapists delivering therapy, and the independent raters assessing the sessions. A sample of sessions was randomly chosen, including sessions delivered by qualified therapists and by students, and sessions from different time points in the treatment process. The fidelity was rated by the first author, and the fidelity rating calculated as a percentage. Comparisons in fidelity scores for the different variables were drawn using Mann–Whitney tests. The reliability of the checklist was assessed through inter and intra-rater reliability testing, and the results were analysed using Kappa statistics.

Outcomes & Results: High fidelity was found across all therapy conditions with a mean score of 98.2%. Fidelity scores were not affected by the administrator of therapy; sessions delivered by qualified and student therapists were rated equally highly. There was a small but significant effect of time, with later treatment

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *International Journal of Language & Communication Disorders* published by John Wiley & Sons Ltd on behalf of Royal College of Speech and Language Therapists

sessions scoring more highly than earlier sessions. However, scores across both periods > 90%. Inter-rater reliability found a high percentage agreement of 93.3% and a Poor Kappa agreement level. Intra-rater agreement found a high percentage agreement of 97.3% and a Fair Kappa agreement level.

Conclusions & Implications: The CommuniCATE reading therapy was implemented as per the protocol across time points, and withstood delegation to students. The high fidelity and good reliability scores have positive implications for the study's validity and reliability, and for the study's replication.

What this paper adds

What is already known on the subject

- TF refers to the degree to which the delivery of core components of a treatment matches the implementation guidelines, that is, the adherence to protocol. Despite the acknowledged importance of TF reporting, this is often neglected in the literature.

What this paper adds to existing knowledge

- This paper shows that the TF assessment of the CommuniCATE study (reading strand) found a 98.2% fidelity score, and that high fidelity was not compromised across treatment conditions. This paper outlines the principles of TF and highlights the need for measures to be in place to establish TF, for example, manuals, training and supervision; and to monitor TF, for example, via the use of checklists. This paper also underlines the scarcity of TF measures and checks in aphasia research. This paper therefore serves as a model of TF practice in aphasia therapy research.

What are the potential or actual clinical implications of this work?

- This study contributes to the findings of the CommuniCATE project (reading strand), and the high fidelity findings enhance the validity of the project and indicate that the therapy manual and training enable accurate implementation of delivery. This paper also contributes to the literature on TF evaluation in aphasia studies, which is presently lacking, and highlights the need for increased focus on the optimum strategies of TF reporting.

INTRODUCTION

Background

Treatment fidelity (TF) refers to the degree to which the delivery of core components of a treatment matches the delivery implementation guidelines; that is, the therapist's adherence to protocol (Kladouchou et al. 2017). Different terms are used in the literature to describe TF, such as procedural reliability, implementation fidelity, treatment integrity and procedural fidelity (Hayden et al. 2015,

Schlosser 2002). In this paper, the term 'treatment fidelity' will be used.

The rationale for ensuring TF

Ensuring TF is critical when designing and evaluating new treatments (Bellg et al. 2004, Hennessey and Rumrill 2003, Moncher and Prinz 1991). We cannot appraise treatment outcomes unless we are confident that the treatment, as described, has been administered. If two therapies have

been compared, it is again essential to show that each has followed protocol. If this is not checked, the comparison may be invalidated, for example, because the treatments may have drifted into overlapping territory. Demonstrating TF is also essential with respect to replication; that is, we need to show that treatment administration adheres to protocol before we can be confident that the same treatment can be delivered by different therapists or in different settings. This requirement is particularly relevant for speech and language therapy, where the delivery of interventions may be delegated, for example, to support staff (Wenke et al. 2014).

Failure to evaluate TF during intervention research increases the risk of many types of error (see the arguments in Carroll et al. 2007, Hinkley and Douglas 2013, Steckler and Linnan 2002). The first is a Type I error. This is where the null hypothesis is wrongly rejected. In the context of an intervention study, the null hypothesis states that there is no difference between the experimental condition (e.g. the receipt of a novel therapy) and control condition (e.g. usual care). In rejecting this, the researcher concludes that there is indeed a difference, and depending on the direction of the results, may argue for the positive benefits of the novel treatment. However, if TF has not been checked, this conclusion may be invalid, since we cannot be sure that the treatment as described was responsible for these benefits. A Type II error is where the null hypothesis is wrongly accepted. Against the example above, this is where results show no difference between the conditions, leading to the conclusion that the novel treatment is ineffective, or at least no more beneficial than usual care. However, again if TF was not evaluated, we cannot be sure that this is the case. For example, it may be that the poor outcomes arose because the novel treatment was not properly administered. A Type III error is where the null hypothesis is rightly rejected, but for the wrong reasons. In our example, this is where there is a difference between the experimental conditions that can be attributed to the content of the novel therapy. However, without checking TF, we cannot be sure of this conclusion. For example, additional elements may have been introduced to the novel treatment that were beyond the protocol and responsible for the enhanced outcomes.

In line with the arguments above, checking TF is critical for the development of treatment evidence and, by extension, for evidence-based practice (EBP). EBP involves selecting interventions that are supported by the highest available evidence. When applying EBP, the clinician must be confident that such evidence is based on the faithful implementation of the treatment and not extraneous effects (see the arguments in Kaderavek and Justice 2010, Kladochou et al. 2017). Monitoring TF can also enable us to address questions that go beyond whether or not a ther-

apy works. For example, it can help to identify the most active components of treatment, and those which can be most faithfully replicated (McCormack et al. 2017). For reasons such as these, TF is crucial to the evaluation of intervention efficacy.

The assessment of TF may promote the uptake of treatment for a further reason. The degree to which treatments are typically described in the literature has been criticized (Hoffmann et al. 2014), with the argument that under-specification hinders replication. This will be less true if TF has been assessed, since this requires treatment to be delineated. Such a treatment, therefore, should be more readily translated into clinical practice (Hinckley and Douglas 2013).

The concept of TF

The literature lacks consensus about how TF should be conceptualized and assessed (Masterson-Algar et al. 2014). An early, comprehensive framework was developed by Bellg et al. (2004). This framework identifies five areas in which researchers can implement strategies aiming to promote TF. The first is study design. Here, researchers must ensure that the therapy content reflects the underlying theoretical rationale. Practical measures are needed to ensure that all participants receive the same therapy dose, within and across treatment conditions, and to manage potential setbacks, such as staff leaving.

The second area is training providers. This aims to ensure that all those who are delivering the therapy have the necessary skills, with a view to minimizing drift or deviation from the protocol. Strategies include the standardization of training, for example, via a training manual; assessment of skill acquisition, for example, by testing or observing trained staff; and provision of post-training booster or supervision sessions, to refresh and maintain the acquired skills. The authors also make the point that training should accommodate staff differences, for example, by increasing the support given to inexperienced providers.

The third area is treatment delivery. Many strategies are recommended here to ensure that treatment is delivered as intended, including the use of a manual, or scripted protocol, and measures to prevent contamination if treatments are being compared. A key strategy involves checking whether the treatment components have been delivered. Such checking may be conducted by the provider. However, the risk of bias is reduced by using an external assessor, for example, scoring a random selection of recorded sessions. The remaining TF categories shift the focus to the person receiving therapy. One is receipt of treatment. This aims to ensure that the patient can

understand and comply with the therapy. Strategies include the provision of accessible information, monitoring of skills during therapy sessions, review of homework tasks and eliciting feedback from patients. The final category is treatment enactment. This concerns the ability of the patient to apply the skills learnt in therapy to real life settings.

The Consolidated Framework for Implementation Fidelity (Carroll et al. 2007) further specifies the components relating to treatment delivery. According to this framework, fidelity is high if an implemented intervention adheres completely to the content, frequency, duration and coverage prescribed by the designers. Content refers to the active components of therapy, such as treatment tasks and strategies. Frequency, duration and coverage refer broadly to the 'dose', for example, the number and length of sessions and whether all participants received the prescribed amount of treatment. According to this framework, implementation fidelity is influenced by four moderating factors. The key moderating factor here is the complexity of intervention, with the authors making the point that adherence is easier to achieve with simple rather than complex treatments. This point is germane to stroke rehabilitation, where interventions are typically complex in nature (Masterson-Algar et al. 2014), and therefore should be evaluated accordingly in terms of their adherence (see the arguments in Hawe et al. 2004). For example, many aphasia therapies, such as that explored in this study, involve multiple components and require adaptation for individual needs (Caute et al. 2019). The second moderator is the application of facilitation strategies. These echo many of the strategies advocated by Bellg et al. (2004), such as the use of treatment manuals and training for intervention providers. The remaining moderators are quality of delivery and participant responsiveness. Quality of delivery is referred to as the manner in which the therapist delivers the intervention appropriately for the individual (Breitenstein et al. 2010, Mihalic 2004). This is deemed important for TF because a treatment may be delivered faithfully in terms of adherence to protocol but the delivery may be poor in quality (Kaderavek and Justice 2010). Participant responsiveness focuses on the recipients of therapy. Here, a lack of commitment or compliance may threaten fidelity.

Assessment and reporting of TF

The above frameworks identify factors that should be included when reporting TF. Studies should indicate the degree to which strategies such as those advocated by Bellg et al. (2004) have been deployed in an attempt to secure fidelity, for example, with respect to training of

providers, specification of therapy and delivery of the treatment dose. Steps taken to secure comprehension of, and compliance with, the intervention are also relevant to treatment receipt.

Whether or not treatment delivery adheres to protocol should be directly assessed. Bellg et al. (2004) reported that the gold standard of TF assessment involves direct or video observation of treatment sessions that are scored against a checklist by a trained coder, with a priori criteria to evaluate therapy delivery according to protocol. There is disagreement on the amount of data required for adequate TF measurements (Behn 2016). Hinckley and Douglas (2013) found that, in aphasia research most of the studies assessed 10–20% of data for TF. Heilemann et al. (2014) recommended 15–40% should be checked for TF measurements. A total of 15% of all treatment sessions were assessed in the current study.

Following the framework of Carroll et al. (2007), the TF checklist needs to reflect the core content of therapy, or active ingredients. The tool deployed in the current study was designed to be multidimensional (Heilemann et al. 2014) and cover procedural adherence to prescribed protocol (Hayden et al. 2015). In line with arguments presented by Carroll et al. (2007), it also attempted to capture elements of delivery quality or the therapist's ability to deliver the therapy appropriately (Breitenstein et al. 2010).

The effectiveness of the checklist depends upon the psychometric soundness of the measure (Stein et al. 2007). Discussion with those who developed the intervention and piloting of the checklist should be deployed to promote validity. Inter and intra-rater reliability of checks should also be assessed, and subject to statistical analysis (Kladouchou et al. 2017). Procedures to eliminate bias include the random selection of sessions to be rated, and the use of trained, independent raters. Bellg et al. (2004) suggest that adherence may decline as a study progresses. TF checking, therefore, should be applied to sessions from different time points in a study. Therapist differences can also be explored by sampling sessions from different providers. It should also be noted that fidelity to a treatment protocol is not binary and therefore the degree to which fidelity is found needs to be quantified on a continuum scale (Moncher and Prinz 1991).

TF in the literature

Despite the acknowledged importance of TF promotion and reporting, it is often neglected in the EBP literature (Kaderavek and Justice 2010). Borrelli et al. (2005) assessed the TF reporting in health behaviour research across 10 years, between 1990 and 2000, and found that only 35% of 342 articles evaluated reported use of a treatment

manual, 22% provided supervision and 27% checked treatment delivery adherence to the protocol.

Looking more specifically at TF reporting in aphasia treatment studies, the studies that do report TF are often not clear about the methods used, for example, observation method, scale-type used, or inter and intra-rater reliability checking methods (Kladouchou et al. 2017). Limits on time and financial resources are both possible reasons for poor TF reporting (McCormack et al. 2017). Hinckley and Douglas (2013) undertook a study of the reporting of TF in aphasia studies across three journals from 2002 to 2011. They found that only 14% ($n=21$) of the 149 studies explicitly reported some aspect of TF, and only 13 of those 21 had one or more raters checking adherence to protocol using videotapes from a sample (between 10% and 20% of the therapy sessions) and calculating a percentage fidelity. Where more than one rater was used, point-to-point agreement between raters was reported. Hinckley and Douglas also found that 2 of the 21 studies listed use of a training manual as a way to implement TF, and 1 used training. Of the 21, only 1 study (Hickey et al. 2004) used two forms of TF measure (a manual and TF ratings).

More recently, Brogan et al. (2019) provided an update on the findings of Hinckley and Douglas (2013) by looking at aphasia treatment studies published from 2012 to 2017; although their review was confined to randomized controlled trials only. Their key finding was that only 21% of reviewed studies ($n=9$) explicitly reported on TF processes. When components of the Bellg et al. (2004) framework were explored, the most commonly reported area was dosage, while the most neglected related to 'participant use of behavioural skills', or enactment.

The lead author of this study explored TF reporting since Brogan et al.'s (2019) findings. The EBSCOhost databases were searched for peer-reviewed papers published since 2017, using the following terms:

AB aphasia

AND

AB 'treatment fidelity' OR AB 'treatment integrity' OR AB 'implementation fidelity' OR AB 'procedural integrity'

A title and abstract review was conducted to identify relevance (Maddy et al. 2014), and the full body of the paper was used for clarification on methods of TF measurement used. The studies generated by these searches were then searched for additional referenced relevant literature.

Excluding Brogan et al. (2019), 12 studies published since (and including) 2017 were found using the above search terms. Of the 12 papers generated, 6 were aphasia intervention studies which reported also on the TF of the study as part of the paper (Cherney et al. 2019, Hilari et al. 2019, Marshall et al. 2018, 2020, Shrubsole et al. 2018, Volkmer et al. 2018). 5 were looking solely at the TF reporting of a specific aphasia intervention, as this paper intends to do (Ball et al. 2018, Carragher et al. 2019, Conlon et al. 2020, Kladouchou et al. 2017, Spell et al. 2020). It should be noted that 3 were study protocol papers, and were therefore giving only intended TF reporting processes (Carragher et al. 2019, Hilari et al. 2019, Volkmer et al. 2018). The final study of the 12 was a review of verb therapies, with a focus on TF reporting (Hickin et al. 2020.)

Of the 12 papers looking at TF, reporting on dosage was universal, but usually in relation to the *intended* dosage. More subtle aspects of dose, such as levels of practice on target items, were neglected. 7 papers explicitly addressed use of a manual to improve TF (Carragher et al. 2019, Conlon et al. 2020, Hilari et al. 2019, Kladouchou et al. 2017, Marshall et al. 2020, Spell et al. 2020, Volkmer et al. 2018). 9 reported use, or intended use, of videos to record the intervention and assess TF (Ball et al. 2018, Carragher et al. 2019, Conlon et al. 2020, Hilari et al. 2019, Kladouchou et al. 2017, Marshall et al. 2018, 2020, Spell et al. 2020, Volkmer et al. 2018). 8 explicitly addressed the use of a checklist (Carragher et al. 2019, Conlon et al. 2020, Hilari et al. 2019, Kladouchou et al. 2017, Marshall et al. 2018, 2020, Shrubsole et al. 2018, Volkmer et al. 2018). 6 of the studies that reported on TF also assessed, or planned to assess, the reliability of their results (Carragher et al. 2019, Hilari et al. 2019, Kladouchou et al. 2017, Marshall et al. 2020, Spell et al. 2020, Volkmer et al. 2018).

Aims of the current study

This study explores the fidelity of treatment delivery in the reading wing of the CommuniCATE project (Caute et al. 2019). The research questions were:

- What is the degree to which the planned components of the CommuniCATE therapy programme were delivered and adhered to?
- Is there a significant difference in adherence to protocol between qualified therapists and students?
- Does TF vary over time; do sessions delivered early or later in the treatment process differ in their TF scores?
- Is the TF evaluation tool used reliable?

It was hypothesized that, with the measures put in place during the delivery of CommuniCATE, high fidelity would

be found. It might be predicted that delegation of therapy to unqualified staff would induce drift from the protocol. However, this was not hypothesized here because the students were trained, closely supervised, and guided by a treatment manual. Alternative possible hypotheses were formulated with respect to time. Early sessions may be vulnerable to drift because the therapy has not been fully mastered; alternatively, this may be a greater risk in later sessions because familiarity encourages less reference to the manual.

It was hoped that the checklist tool would be found to be reliable thanks to the measures taken to ensure it captured the aims of the therapy and the piloting.

METHODOLOGY

This research was part of the reading wing of the CommuniCATE study (Caute et al. 2019) which explored the benefits of technology-enhanced reading therapy for people with aphasia.

Participants

Participants in the CommuniCATE study were People with Aphasia (PwA), qualified Speech and Language Therapists (SLTs) and Students. The 21 clients who received reading therapy were recruited for CommuniCATE; all had aphasia following a stroke that had occurred at least 4 months before their recruitment, were medically stable and did not have severely impaired cognition. All participants gave informed written consent. The three qualified SLTs were employees of City, University of London and were leading the CommuniCATE project and delivering therapy. The students who delivered the therapy under the supervision of the therapists were all students at City, University of London, studying for a Bachelor's or Master's in Speech and Language Therapy.

Ethical approval

The CommuniCATE study received ethical approval from the Bromley (London) NRES Committee (14/LO/1531) and from the School of Health Sciences Research Ethics Committee, City, University of London.

Intervention

The CommuniCATE reading therapy was delivered face to face and was prescribed in a manual (see [appendix](#)

B). Therapy sessions provided training in the technology (Claro Software and Amazon e-readers) and offered structured reading practice using the technology. The CommuniCATE therapy is a complex intervention. That is, a treatment which is not highly prescribed, is often individualized as appropriate per client and contains several key components which means the treatment is not easily definable (Behn 2016).

Fidelity strategies

The following steps were taken by the CommuniCATE project to promote TF.

Study design: The study employed a randomized controlled design comparing an immediately treated group with a waitlist control group. Controls received the prescribed therapy after a midpoint cross-over. Those randomized to the control group had no access to the therapy technologies or materials before cross-over, and screening at recruitment ensured that participants were not receiving speech and language therapy elsewhere. The treatment dose (1 session for technology set up, followed by 12 hours of individual therapy, delivered over 6 weeks) was specified in advance and receipt was individually monitored. Thus, the design ensured that the same treatment dose was received within conditions, and minimized contamination. Treatment was delivered by a team of qualified and student clinicians, ensuring that there were contingencies for setbacks, such as staff absence.

Training of providers: All student clinicians received a standardized, 1 day training course on the rationale, content and delivery of therapy. The training was specified in a manual and used consistent materials, such as PowerPoint slides. Skill acquisition was checked via exercises and discussion. Weekly supervision and regular observation of sessions followed by individual feedback aimed to minimize drift and take account of provider differences.

Treatment delivery: Treatment was specified in a manual and supported by standardized materials, such as instructions on how to use the assistive technologies that were employed in therapy. Weekly supervision of providers and observation of sessions monitored adherence.

Receipt of treatment: Participant comprehension of and engagement with therapy was promoted through aphasia-friendly materials, for example, to support goal-setting, training in technology use and conduct of reading exercises. Supportive communication techniques, such as Talking Mats (Murphy 1998), were employed to ensure that participants were able to follow instructions and contribute their ideas. Therapy included regular monitoring and discussion of progress, including execution of reading tasks to check skill acquisition.

Treatment enactment: Independent reading was promoted through between-session reading tasks. These were recorded in a diary that was reviewed each week.

Adherence to protocol

This study seeks to assess the adherence of treatment to the CommuniCATE therapy protocol using retrospective video recordings and a tool designed to assess fidelity.

Research team

Two independent raters assessed the fidelity of the therapy sessions (Rater 1 - first author; and Rater 2 - Author CM). They were both qualified SLTs, but not part of the CommuniCATE team involved in the delivery of the reading therapy. Rater 2 was familiarized with the checklist by Rater 1 and trained in its use by applying it to a filmed treatment session. This training session was not included in the final data set.

Data

All therapy sessions ($n=252$) were filmed. Videos of 15% ($n = 38$ sessions) featuring 19/21 clients were analysed retrospectively. Thus, this study extends the TF data reported in Caute et al. (2019), which was based on 24 assessed videos. Videos were chosen quasi-randomly from a list of participants in the Reading strand of CommuniCATE. That is, participants were picked at random from a list by the first author who was not part of the treating team and had no access to participant data. Two videos were selected per participant according to their time point: one video of a therapy session occurring early in the therapy process between sessions 2 and 4, and one video of a therapy session occurring towards the end of the therapy process between sessions 8 and 11. These were checked to ensure that there was some mix across all sessions of those delivered by therapists and those delivered by students, or both.

Checklist

In order to assess adherence to protocol, a checklist was developed. This was a list of criteria to which the delivery of the sessions should adhere, and against which the sessions were rated. The checklist was devised by the first author (Rater 1), referring to the CommuniCATE manual, and The Checklist Development Checklist written by Stufflebeam (2000). The intention was for the key components of the therapy to be captured, in terms of its aims, such that, if

a component had not been optimally present in therapy, the treatment's success may have been compromised. The checklist was created for use by raters familiar with the intervention but independent from it.

CommuniCATE is a novel therapy, and the essential components, that is, the active ingredients expected to create therapeutic change, are not known. Therefore, each perceived key component of the therapy programme was assessed (Heilemann et al. 2014). The checklist covers the therapist-critical components of the intervention, that is, the tasks and skills of the therapist that made up the session. The author made judicious use of the CommuniCATE manual and observation of a small number of sample videos (not included in the rating process) to help with creation of the fidelity checklist. Another study that looked at assessing the TF of a complex intervention was also referred to for guidance (Behn 2016). The draft checklist was then submitted to the CommuniCATE therapists twice for review and discussion to ensure content validity (Kladouchou et al. 2017). Amendments, deletions or additions were subsequently made. For example, a component focusing on collaborative decision-making was amended to focus on the collaborative reviewing of progress in line with client goals. The checklist was also piloted with three sessions rated by Rater 1 and 2, which allowed for assessment of its appropriateness and effectiveness. The resulting checklist contained nine components to be assessed for their presence in therapy (see [appendix A](#)).

The checklist for this study aimed to integrate both adherence to procedural elements and delivery quality; the latter being more difficult to assess than the former (Kaderavek and Justice 2010). An example of a component from the checklist designed to assess delivery quality is 'Therapist gives differential feedback', where both the existence of feedback and whether or not it was differentiated to the client's needs had to be assessed.

The checklist employs a Likert-type scale for rating, which awarded a score of 0 (absent), 0.5 (present but could have benefitted from more) or 1 (present to an appropriate extent). There was also an option of 'Not applicable' for each component. This was provided for instances in which a component was judged to be irrelevant for the assessed session. In fact, this option was never used by the primary rater. This Likert-type scale was deemed more appropriate for the rating of the videos than simple present/absent binary rating scales so that the degree of TF could be reported.

Data collection

TF was measured using direct means. Videos of the selected sessions were watched in their entirety and rated.

Inter-rater reliability

To assess reliability of the checklists, an independent second rater, blind to the scores of Rater 1, observed and rated 25 (10%) of the videos. These were randomly selected from the original set watched by Rater 1. Comparisons between the scores of Rater 1 and 2 were drawn.

Intra-rater reliability

To review intra-rater reliability, 10% of the videos were re-rated by Rater 1 three weeks following the original rating, as in Kladouchou et al. (2017). The videos re-watched by Rater 1 for intra-rater reliability were the same that were watched by Rater 2 for inter-rater reliability. The scores given by Rater 2 had not been seen by Rater 1 when they watched the videos for the second time.

Data analysis

From the scores given on the checklist, a percentage TF was calculated per session. Guidelines by Heilemann et al. (2014) based on a literature review were used to ascertain the relative rating of TF. Therefore, a TF percentage of $\geq 80\%$ was considered to be high. Once the results were collected, they were checked for their distribution, using Shapiro Wilk. As all data were not normally distributed, non-parametric tests (Mann-Whitney) were conducted to compare conditions across providers and time.

For the calculation of inter and intra-rater reliability, percentage agreement across scores was calculated, and Kappa coefficient statistics were used, which takes into account the possibility of agreement occurring by chance. A Kappa coefficient of 0.75–1.00 is excellent, 0.60–0.74 is good, 0.40–0.59 is fair and < 0.40 is poor (Fleiss 1981). With respect to percentage agreement, guidelines specify a benchmark of 70% inter-rater agreement to be the acceptable minimum (Heilemann et al. 2014). No guidelines were found for acceptable intra-rater agreement. All analyses were carried out on IBM SPSS v.23.

RESULTS

Treatment fidelity

The number of therapy components rated per session was nine and each session was given an overall fidelity score out of nine, which was converted into a percentage. The average TF score (Rater 1 rating 38 videos across all nine rated components across all sessions) was 8.84/9 (SD =

0.26), which is 98.2%, with scores ranging from 8/9 to 9/9. This demonstrated that therapy was delivered with a high degree of fidelity. Component 2 on the checklist ('therapist involves participants in collaborative reviewing of goals/progress') was most frequently marked lower than 100% fidelity.

The TF scores for sessions delivered by students ($n=15$) were compared with those delivered by qualified therapists ($n=16$), and those delivered jointly by therapists with students ($n=7$). Table 1 shows the total TF score per session given for sessions delivered by students, and the total TF scores given for sessions delivered by therapists. Table 1 also shows the total TF score per session given for sessions delivered jointly by therapists and students. The mean (SD) TF score for the student delivered sessions was 8.87 (0.30) or 98.5% (median = 9.0). The mean (SD) TF score for therapist delivered sessions was 8.84 (0.24) or 98.25% (median = 9.0). The mean (SD) for sessions delivered by both a therapist and student was 8.79 (0.27) or 97.6% (median = 9.0).

Mann-Whitney tests indicate that there was no significant difference between the TF scores for student versus therapist delivered sessions ($U = 109.00$, $p = 0.568$); for student delivered versus jointly delivered sessions ($U = 42.0$, $p = 0.341$); or for therapist versus jointly delivered sessions ($U = 49.5$, $p = 0.599$).

TF scores for early sessions versus late sessions

The TF scores for sessions earlier on in the therapy process ($n=19$) were compared with those delivered at a later stage of the therapy process ($n=19$). Table 1 shows the total TF score per session given for sessions delivered earlier on in the therapy process (denoted with an 'a') and the total TF scores given for sessions later in the therapy process (denoted with a 'b'). The mean (SD) TF score for the early sessions was 8.74 (0.31) or 97.1% (median = 9.0), whereas the mean (SD) TF score for the later sessions was 8.95 (0.16) or 99.4% (median = 9.0). Mann-Whitney tests indicated that there was a significant difference between the ranked scores for the early and late sessions ($U = 113.0$, $p = 0.013$).

Inter-rater reliability

Table 2 shows the TF percentage score from raters 1 and 2 per session, and, below this, the inter-rater reliabilities generated across the 25 sessions that were rated by both Rater 1 and 2.

Comparisons were drawn between scores across the nine components across the 25 sessions ($n=225$ components in total) rated by both Rater 1 and 2, and 93.3%

TABLE 1 Rater 1 fidelity scores across all nine components for all 38 watched videos

| Component | V1a (B) | V1b (S) | V2a (B) | V2b (B) | V3a (T) | V3b (T) | V4a (T) | V4b (T) | V5a (T) | V5b (S) | V6a (S) | V6b (S) | V7a (S) | V7b (S) |
|-----------|-------------|----------|-------------|----------|-------------|----------|-------------|-------------|----------|----------|-------------|----------|----------|----------|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 0.5 | 1 | 0.5 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.5 | 1 | 1 | 1 |
| 7 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Total | 8.5 (94.4%) | 9 (100%) | 8.5 (94.4%) | 9 (100%) | 8.5 (94.4%) | 9 (100%) | 8.5 (94.4%) | 8.5 (94.4%) | 9 (100%) | 9 (100%) | 8.5 (94.4%) | 9 (100%) | 9 (100%) | 9 (100%) |
| Component | V8a (S) | V8b (S) | V9a (T) | V9b (S) | V10a (B) | V10b (S) | V11a (S) | V11b (S) | V12a (B) | V12b (B) | | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 2 | 0.5 | 1 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | | | | |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 7 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | |
| Total | 8.5 (94.4%) | 9 (100%) | 9 (100%) | 9 (100%) | 8.5 (94.4%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | | | | |
| Component | V13a (T) | V13b (S) | V14a (S) | V14b (T) | V15a (T) | V15b (T) | V16a (T) | V16b (T) | V17a (T) | V17b (T) | V18a (T) | V18b (T) | V19a (S) | V19b (B) |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 0.5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Total | 9 (100%) | 9 (100%) | 8 (88.8%) | 9 (100%) | 8.5 (94.4%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) | 9 (100%) |

Note: Sessions a = early; and b = late. TF scores for student delivered versus therapist delivered, versus sessions delivered by both (S = student delivered, T = Therapist delivered, B = delivered by both).

TABLE 2 TF scores per session given by Rater 1 and 2 at time 1 (inter-rater), and by Rater 1 at Time 2 (intra-rater)

| Session | TF % R1 | TF % R2 | TF % R1 T2 |
|---------------------------------------------|----------------------------------|---------|------------|
| V1a | 94.40% | 88.89% | 94.40% |
| V1b | 100% | 100% | 100% |
| V2a | 94.40% | 88.33% | 88.89% |
| V2b | 100% | 100% | 100% |
| V3a | 94.40% | 100% | 94.40% |
| V3b | 100% | 100% | 100% |
| V5a | 100% | 100% | 100% |
| V5b | 100% | 100% | 100% |
| V6a | 94.4% | 100% | 100% |
| V6b | 100% | 88.8% | 100% |
| V7a | 100% | 100% | 100% |
| V7b | 100% | 100% | 100% |
| V8a | 94.4% | 94.4% | 100% |
| V8b | 100% | 100% | 100% |
| V10a | 94.4% | 88.8% | 100% |
| V10b | 100% | 100% | 100% |
| V11a | 100% | 100% | 100% |
| V11b | 100% | 100% | 100% |
| V13a | 100% | 100% | 100% |
| V13b | 100% | 100% | 100% |
| V14a | 88.8% | 100% | 100% |
| V14b | 100% | 88.8% | 100% |
| V18a | 100% | 100% | 100% |
| V18b | 100% | 100% | 100% |
| V19a | 100% | 94.40% | 100% |
| Average inter-rater reliability Kappa value | Level of agreement (Fleiss 1981) | | |
| 0.139 | Poor | | |
| Average intra-rater reliability Kappa value | Level of agreement (Fleiss 1981) | | |
| 0.488 | Fair | | |

percentage agreement was found. Despite this high percentage agreement, the Kappa value was only 0.139 ($p < 0.05$), indicating a Poor level of agreement between the two raters.

Intra-rater reliability

Table 2 shows the TF percentage score per session, given by Rater 1 at Time 2, and below, the intra-rater reliabilities generated across the 25 sessions that were rated by Rater 1 at Time 1 and 2. Comparisons were drawn between scores across the nine components across the 25 sessions ($n=225$ components in total) rated by Rater 1 at Time 1 and 2, and 97.3% percentage agreement was found. The average Kappa value was 0.488 ($p < 0.05$), indicating a Fair level of agreement between Time 1 and 2.

DISCUSSION

This discussion will review the outcomes against the aims of this paper. It will appraise the implications of the findings and if the methodology was sensitive to poor TF. It will acknowledge the limitations of the study and look at the implications for further research.

Treatment fidelity

The average TF score across all sessions rated by rater 1 was 8.84/9 (98.2%), with scores ranging from 8/9 to 9/9, demonstrating that there was a high degree of fidelity in delivery of therapy. Other studies (Kladouchou et al. 2017) report similarly high TF values. It is interesting to note this trend. The reason for this could be sound measures in

place to enhance fidelity of delivery in studies. For example, use of manuals, training and feedback (Carroll et al. 2007). It could also be that, as with CommuniCATE, TF assessments are carried out at the research stage of a therapy and are therefore more likely to be faithful to protocol. It also could be indicative of publication bias. It would be useful, and necessary, to assess the TF of the therapy in other contexts.

TF scores may be influenced by the nature of the component. As observed in the Results section, component 2 on the checklist ('therapist involves participants in collaborative reviewing of goals/progress') was most frequently marked at < 100% fidelity. The nature of this component is such that it may have been subtly entrenched into the threads of the session and not as explicitly marked as to obtain higher marks for fidelity.

In comparisons drawn, there was no significant difference in the TF scores for sessions delivered by students or therapists (or both). This shows that therapist drift was not induced by delegation to unqualified practitioners, possibly because of the manual and provision of training and supervision.

There was a significant effect of time on TF, with late sessions scoring more highly than early sessions. This may suggest that practitioners became more familiarized with the protocol as treatment progressed. However, despite the significant finding, the difference in fidelity scores for early and late sessions was small and, at both time points, TF scores were > 90%. Drift, therefore, was minimal at all stages of the treatment process.

Checklist reliability

This study examined both inter and intra-rater reliability ratings of the fidelity tool, which is rare in aphasia studies in the literature thus far, particularly with respect to intra-rater agreement (see discussions in Heilmann et al. 2014, Kladouchou et al. 2017). Heilmann et al. (2014) reported $\geq 70\%$ to be a good level of inter-rater agreement. Our results exceeded this level by a considerable margin with 93.3% for inter-rater agreement, and 97.3% for intra-rater agreement. Despite these high percentages, the Kappa results were disappointing. The Kappa statistic yielded a Poor value for inter-rater agreement, and only a Fair value for intra-rater agreement. Raters had to choose between four options on the rating scale, but selected (1) 'present to an appropriate extent' on the vast majority of occasions. This meant that the chance of disagreement between raters was very low. As the Kappa statistic evaluates achieved agreement against chance, the value was depressed. The few disagreements between the raters were most common on component nine ('therapist gives differential feedback'), with

scores varying between (1) and (0.5). Thus, it was interesting that a judgement that included treatment quality proved to be the most problematic.

As in other protocols, 'Not applicable' was given as an option, pertaining to instances in which a component was deemed absent or redundant, but that this did not compromise the aims of the therapy because of the varying individual needs of each client. This might be used, for example, for component 4 ('Therapist uses appropriate tools to scaffold and support the session, where necessary'). However, 'Not applicable' was only used, rarely, by rater 2, which highlights that all nine of the components on the checklist were not transient elements but were threaded throughout the entire therapy process.

Other TF studies that have checked inter-rater reliability tend to report high values (Kladouchou et al. 2017). It should be noted though that some studies report inter-rater reliability using point-to-point agreement rather than using statistical coefficients such as Kappa. This method does not account for chance agreements (Hayden et al. 2015) and so findings should be interpreted cautiously (Kladouchou et al. 2017).

In terms of data used for this study, guidelines as per Heilemann et al. (2014) were that a minimum of 15% of sessions should be rated for TF and 10–30% be checked by a secondary rater. The results obtained here showed very little variation, suggesting that watching further sessions would still not have affected the findings. In Cauter et al. (2019), only 24 videos were checked. The larger sample in this study still yielded very high TF values, but it did expose a subtle time effect. This may indicate that fidelity improved over time. However, the difference was small, pointing to a possible type 1 error.

Limitations

The direct method used to assess TF in this study, that is, videos watched retrospectively by independent raters and evaluated using an objective measuring tool, is considered the gold standard in TF assessment, giving objective and detailed data (Kladouchou et al. 2017). Although this was a strength of the study, it is important to acknowledge the limitations.

This study looks only at the TF of the CommuniCATE project as regards the therapy implementation, to include content and delivery quality. However, TF is also improved through standardized training of treatment providers (Adams et al. 2012), use of scripted treatment manuals, supervisory feedback and measures to support treatment receipt (Hinckley and Douglas 2013). These additional factors were employed in the current study, but not assessed. For example, training of the treating students

was not observed and/or scored. Additionally, self-rating of perceived fidelity scores completed by the treating therapists and students could have been included.

This study also did not look at other moderating factors impacting the degree of TF, as reported by Carroll et al. (2007), such as participant responsiveness, dosage or treatment receipt or enactment (Bellg et al. 2004). For example, the length of the sessions assessed for this study tended to vary slightly which may be useful to evaluate. The prescribed length per sessions was 1 hour; however, several of the sessions ran over this.

These factors could be looked at in further studies, as could examining therapist views on the quality of the therapy manual, as done by Kladouchou et al. (2017). Qualitative research methods, such as questionnaires and interviews, could be useful methods for obtaining these data and may augment interpretation of findings (Kladouchou et al. 2017), as could use of secondary data sources to inspect TF, for example, therapy session notes (Heilemann et al. 2014). Other, non-therapist critical components which arose in sessions, for example, failures of the technology, could also have been looked at.

A further limitation is that, although what were considered to be the core components of the therapy were all included in the checklist as recommended (Carroll et al. 2007), the active ingredients most responsible for change were not known. This means that opportunity to draw conclusions about the relative importance of each active ingredient and its fidelity is lost and that the TF checklists were perhaps not as nuanced as they could be (Kladouchou et al. 2017).

Identifying agents for positive change is a big challenge for researchers, particularly those looking at complex interventions, despite this being essential for the establishment of TF (Behn 2016). Complex interventions, as defined by lead thinkers in this area, the Medical Research Council (Craig et al. 2008), are those with multiple key components that may be more difficult to define, particularly non-drug interventions, such as stroke rehabilitation. TF monitoring of complex interventions should include qualitative aspects pertaining to delivery quality and procedural adherence, as well as to the interaction with the context of delivery, for example, societal and cultural systems and organizations (Bragstad et al. 2019), which would create protocols that can be better transferred to non-research settings (McCormack et al. 2017). Conducting sensitivity analyses using TF data and outcomes from studies of the same treatment intervention is one method of identifying the essential components of an intervention (Carroll et al. 2007) versus the adaptable components (Damschroder et al. 2009). When designing a therapy, consideration of the fundamental active ingredients of the intervention is paramount, in order for those key components to be mon-

itored for their presence, for the intervention to be replicated, and for the efficacy and usefulness of the intervention to be evaluated (Walker et al. 2017).

Clinical and research implications

This study contributes to the findings of the CommuniCATE project (reading strand) and the high fidelity findings enhance the validity of the project and indicate that the therapy manual and training enable accurate implementation of delivery. The evaluation process also enhances its replicability beyond research. This paper also contributes to the literature on TF evaluation in aphasia studies, which is presently lacking, and highlights the need for increased focus on the optimum methodological strategies of TF reporting, particularly in aphasia studies (Kladouchou et al. 2017). As Bellg et al. (2004) contend, ignoring TF practices is ultimately more costly in terms of resources.

As an intervention is implemented in clinical practice, it is important that TF continues to be monitored, as it is expected that the fidelity of the implementation will be reduced as the therapist adapts the therapy to respond to individual and contextual needs (Kadaverek and Justice 2010). It is necessary to assess TF when a therapy is used in a clinical rather than research context, to show that the therapy can be delivered faithfully across different contexts.

As the therapist adapts the therapy as appropriate whilst adhering to protocol (Masterson-Algar et al. 2017), the issue of degree of permissible adaptation must be addressed (Craig et al. 2008). The pro-adaptation approach suggests that implementations which deviate and adapt well to local needs are more likely to be successful, as long as fidelity to the identified core ingredients is maintained (Carroll et al. 2007). Further research is needed into the TF assessment practices of complex interventions, such as this one. It is important that an acceptable level of standardization versus adaptation in complex interventions is agreed upon (Hawe et al. 2004). Yet, this is an area in which there is a lack of consensus in the literature (Kladouchou et al. 2017). In line with this, Walker et al. (2017) found that in a survey of 182 stroke rehabilitation trial papers, only 42% of papers reported a clear, full description of the core intervention components. Here, the identification of active ingredients in a therapy protocol is key. In the current study, all nine components on the checklist achieved high fidelity scores, with none scoring < 90%. This may suggest that all components made an active contribution to the positive therapy outcomes reported in our companion paper (Caute et al. 2019). However, the fact that each component was faithfully realized may not mean that each is


necessarily crucial for a positive outcome. Further studies that systematically manipulate the presence (or absence), or treatment components at the design stage (Craig et al. 2008), might further tease this issue apart. Such studies might explore whether any components can be omitted or adapted from the therapy protocol without a detriment to outcomes.

TF has a crucial role to play in developing scientific basis for good EBP, the goal of which is to demonstrate that intervention has positive outcomes and to determine which treatments are most effective for which disorders (Kaderavek and Justice 2010). In making such judgements we must be confident that the treatment has been administered according to the protocol. These interventions may be those on which clinical guidelines for practice are founded (Hinckley and Douglas 2013), and therefore more speech and language therapy studies need to incorporate TF assessment and to be transparent about the methods used to assess this (Kladouchou et al. 2017).

DATA AVAILABILITY STATEMENT

The data are available on request from the authors.

ORCID

Katharine Bacon  <https://orcid.org/0000-0001-8607-2432>

REFERENCES

- Adams, C., Lockton, E., Gaile, J., Earl, G. & Freed, J. (2012) Implementation of a manualized communication intervention for school-aged children with pragmatic and social communication needs in a randomized controlled trial: The Social Communication Intervention Project. *International Journal of Language and Communication Disorders*, 47(3), 245–256.
- Ball, A.L., De Riesthal, M. & Steele, R.D. (2018) Exploring treatment fidelity in persons with aphasia autonomously practicing with computerized therapy materials. *American Journal of Speech-Language Pathology*, 27(1S), 454–463.
- Behn, N. (2016) *Communication and quality of life outcomes in people with acquired brain injury following project-based treatment*. Unpublished Doctoral thesis, City University London. Available at: https://city.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwdV1LSwMxEB4UL2IPihatVuYPbOnG7iZ7VteCCh7Eg5eSxwxUZBdpRbz4253srlqLnhNCZibM48s8AM7UaJys6QQTFOrZ61yzDXnBLjeUjk1BlPqMmzYG00dd3qmmrm-y665qz-IEyRl5c0rUO-iqGNlq072au4uiCj4dbUatx44qBKHdh52LIY3sPNqjah5df1RcoMTu2NYzvWDM-z5mwfl2KyGmB8wrbZG6MyChaHzN0KaCLlxxk9UloRyah1W9ibLADUJjohAJ-Z4sfwLC8vD-fJvGCsw6amX0RUag-9GxMaa-WTelbOARUXnljUhaPwE4ylznxtIzT2ohisGpCR9D_-7DBfwvHsC32vkMQTmCL5YXTsGXaacPBT417g7I (Accessed: 02 August 2018).
- Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M., et al. and Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004) Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, 23(5), 443–451.
- Borrelli, B., Sepinwall, D., Bellg, A.J., Breger, R., Defrancesco, C., Sharp, D.L., et al. (2005) A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting & Clinical Psychology*, 73(5), 852–860.
- Bragstad, L.K., Bronken, B.A., Sveen, U., Hjelle, E., Kitzmuller, G., Martinsen, R., et al. (2019) Implementation fidelity in a complex intervention promoting psychosocial well-being following stroke: An explanatory sequential mixed methods study. *BMC Medical Research Methodology*, 19(59), 1–18.
- Breitenstein, S.M., Gross, D., Garvey, C.A., Hill, C., Fogg, L. & Resnick, B. (2010) Implementation fidelity in community-based interventions. *Research in Nursing & Health*, 33(2), 164–173.
- Brogan, E., Ciccone, N. & Godecke, E. (2019) Treatment Fidelity in aphasia randomised controlled trials. *Aphasiology*, 33, 759–779.
- Carragher, M., Ryan, B., Worrall, L., Thomas, S., Rose, M., Simmons-Mackie, N., et al. (2019) Fidelity protocol for the Action Success Knowledge (ASK) trial: A psychosocial intervention administered by speech and language therapists to prevent depression in people with post-stroke aphasia. *BMJ open*, 9(5), e023560.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. & Balain, S. (2007) A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40.
- Caute, A., Woolf, C., Wilson, S., Stokes, C., Monnelly, K., Cruice, M., et al. (2019) Technology-enhanced reading therapy for people with aphasia: findings from a quasi-randomized waitlist controlled study. *Journal of Speech, Language, and Hearing Research*, 62(12), 4382–4416.
- Cherney, L.R., Braun, E.J., Lee, J.B., Kocherginsky, M. & Van Vuuren, S. (2019) Optimising recovery in aphasia: learning following exposure to a single dose of computer-based script training. *International Journal of Speech-Language Pathology*, 21(5), 448–458.
- Cocks, N., Pritchard, M., Cornish, H., Johnson, N. & Cruice, M. (2013) A “novel” reading therapy programme for reading difficulties after a subarachnoid haemorrhage. *Aphasiology*, 27(5), 509–531.
- Conlon, E.L., Braun, E.J., Babbitt, E.M. & Cherney, L.R. (2020) Treatment fidelity procedures for an aphasia intervention within a randomized controlled trial: Design, feasibility, and results. *American Journal of Speech-Language Pathology*, 29(1S), 412–424.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., et al. (2008) Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal*, 337(7676), 979–983.
- Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A. & Lowery, J.C. (2009) Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(1), 50.
- Fleiss, J.L. (1981) *Statistical methods for rates and proportions*. New York: Wiley.
- Hawe, P., Shiell, A. & Riley, T. (2004) Complex interventions: how ‘out of control’ can a randomised controlled trial be? *British Medical Journal*, 328(7455), 1561–1563.
- Hayden, D., Namasivayam, A.K. & Ward, R. (2015) The assessment of fidelity in a motor speech-treatment approach. *Speech, Language and Hearing*, 18(1), 30–38.
- Heilemann, C., Best, W., Johnson, F., Beckley, F., Edwards, S., Maxim, J., et al. (2014) Investigating treatment fidelity in a conversation-

- based aphasia therapy. *Aphasie und Verwandte Gebiete*, 37(2), 14–26.
- Helm- Estabrooks, N. (2001) *Cognitive Linguistic Quick Test*. San Antonio, TX: The Psychological Corporation.
- Hennessey, M.L. & Rumrill, P.D. (2003) Treatment fidelity in rehabilitation research. *Journal of Vocational Rehabilitation*, 19(3), 123–126.
- Hickin, J., Cruice, M. & Dipper, L. (2020) A systematically conducted scoping review of the evidence and fidelity of treatments for verb deficits in aphasia: Verb-in-isolation treatments. *American Journal of Speech–Language Pathology*, 29(1S), 530–559.
- Hickey, E., Bourgeois, M. & Olswang, L. (2004) Effects of training volunteers to converse with nursing home residents with aphasia. *Aphasiology*, 18(5–7), 625–637.
- Hilari, K., Behn, N., Marshall, J., Simpson, A., Thomas, S., Northcott, S., et al. (2019) Adjustment with aphasia after stroke: study protocol for a pilot feasibility randomised controlled trial for Supporting wellbeing through PEer Befriending (SUPERB). *Pilot Feasibility Studies*, 14(5),
- Hinckley, J.J. & Douglas, N.F. (2013) Treatment fidelity: Its importance and reported frequency in aphasia treatment studies. *American Journal of Speech–Language Pathology*, 22(2), S284.
- Hoffmann T, C., Glasziou P, P., Boutron, I., Milne, R., Perera, R., Moher, D., et al. (2014) Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *British Medical Journal*, 348, g1687.
- Kaderavek, J.N. & Justice, L.M. (2010) Fidelity: An essential component of evidence-based practice in speech–language pathology. *American Journal of Speech–Language Pathology*, 19(4), 369–379.
- Kladouchou, V., Papanthanasidou, I., Efstratiadou, E.A., Christaki, V. & Hilari, K. (2017) Treatment integrity of elaborated semantic feature analysis aphasia therapy delivered in individual and group settings. *International Journal of Language & Communication Disorders*, 52(6), 733–749.
- LaPointe, L.L. & Horner, J. (1998) *Reading comprehension battery for aphasia – Second edition. Examiners manual.* Austin: Pro-Ed.
- 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), 254–67.
- Marshall, J., Devane, N., Edmonds, L., Talbot, R., Wilson, S., Woolf, C., et al. (2018) Delivering word retrieval therapies for people with aphasia in a virtual communication environment. *Aphasiology*, 32(9), 1054–1074.
- Marshall, J., Devane, N., Talbot, R., Cauter, A., Cruice, M., Hilari, K., et al. (2020) A randomised trial of social support group intervention for people with aphasia: A Novel application of virtual reality. *Plos One*, 24;15(9), e0239715.
- Masterson-Algar, P., Burton, C.R., Rycroft-Malone, J., Sackley, C.M. & Walker, M.F. (2014) Towards a programme theory for fidelity in the evaluation of complex interventions. *Journal of Evaluation in Clinical Practice*, 20(4), 445–452.
- Masterson-Algar, P., Burton, C.R., Brady, M.C., Nicoll, A., Clarke, C.E., Rick, C., et al. (2017) The PD COMM trial: A protocol for the process evaluation of a randomised trial assessing the effectiveness of two types of SLT for people with Parkinson’s disease. *Trials*, 18(1),
- Mccormack, J., Baker, E., Masso, S., Crowe, K., Mcleod, S., Wren, Y., et al. (2017) Implementation fidelity of a computer-assisted intervention for children with speech sound disorders. *International Journal of Speech–Language Pathology*, 19(3), 265–276.
- Mihalic, S. (2004) The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4, (83–86), 86 and 99–105.
- Moncher, F.J. & Prinz, R.J. (1991) Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11(3), 247–266.
- Murphy, J. (1998) Talking Mats: Speech and language research in practice. *Speech and language therapy in practice. Autumn*, 1998, 11–14.
- Schlosser, R. (2002) On the importance of being earnest about treatment. *Integrity, Augmentative and Alternative Communication*, 18(1), 36–44.
- Shrubsole, K., Worrall, L., Power, E. & O’connor, D.A. (2018) The Acute Aphasia IMplementation Study (AAIMS): A pilot cluster randomized controlled trial. *International Journal of Language & Communication Disorders*, 53(5), 1021–1056.
- Spell, L.A., Richardson, J.D., Basilakos, A., Stark, B.C., Teklehaimanot, A., Hillis, A.E., et al. (2020) Developing, implementing, and improving assessment and treatment fidelity in clinical aphasia research. *American Journal of Speech–Language Pathology*, 29(1), 286–298.
- Stekler, A. & Linnan, L. (2002) *Process evaluation for public health interventions and research* San Francisco, CA: Jossey-Bass.
- Stein, K.F., Sargent, J.T. & Rafaels, N. (2007) Intervention research: Establishing fidelity of the independent variable in nursing clinical trials. *Nursing Research*, 56(1), 54–62.
- Stufflebeam, D.L. (2000) Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist (CDC). Available at: https://www.wmich.edu/sites/default/files/attachments/u350/2014/guidelines_cdc.pdf. (Accessed on 2 August 2018).
- Swinburn, K., Porter, G. & Howard, D. (2005) *The Comprehensive Aphasia Test*. Hove, UK: Psychology Press.
- Volkmer, A., Spector, A., Warren, J.D. et al. (2018) The ‘Better Conversations with Primary Progressive Aphasia (BCPPA)’ program for people with PPA (Primary Progressive Aphasia): protocol for a randomized controlled pilot study. *Pilot Feasibility Studies*, 4, 158.
- Walker, M.F., Hoffmann, T.C., Brady, M.C., Dean, C.M., Eng, J.J., Farrin, A.J., et al. (2017) Improving the development, monitoring and reporting of stroke rehabilitation research: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke: Official Journal of the International Stroke Society*, 12(5), 472–479.
- Wenke, R., Lawrie, M., Hobson, T., Comben, W., Romano, M., Ward, E., et al. (2014) Feasibility and cost analysis of implementing high intensity aphasia clinics within a sub-acute setting. *International Journal of Speech–Language Pathology*, 16, (3), 250–259.
- Wiederholt, J.L. & Bryant, B.R. (2001) *Gray Oral Reading Tests—Fourth Edition*. Austin, TX: Pro-Ed.

How to cite this article: Bacon, K., Marshall, J., Cauter, A., Monnelly, K., Cruice, M., Moutou, C. & Woolf, C. (2021) Treatment fidelity of technology-enhanced reading therapy (CommuniCATE) for people with aphasia. *International Journal of Language & Communication Disorders*. 1–18. <https://doi.org/10.1111/1460-6984.12637>

APPENDIX A: CHECKLIST

AIM of Tx: To investigate the effects of speech and language therapy using computer technologies for people with aphasia, including whether changes in language and communication also improve social participation and quality of life

| Therapy component | Present to an appropriate extent | Present to some degree but could have benefitted from more | Absent | Not applicable |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|------------------------------------------------------------|--------|----------------|
| <i>Activity checklist (at least one of the following activities occurs)</i> | | | | |
| Accessing and reading digital text | | | | |
| Reading paper-based text | | | | |
| Review and/or practice of reading technology | | | | |
| Discussion of reading activities | | | | |
| <i>Facilitation strategies checklist</i> | | | | |
| Therapist involves participants in collaborative reviewing of goals/progress | | | | |
| Therapist facilitates identification of reading problems, if any, and supports/facilitates solving them, where necessary | | | | |
| Therapist uses appropriate tools to scaffold and support the session, where necessary, for example, visuals, writing, repetition | | | | |
| Therapist ensures that the content of the session is accessible (i.e., of interest and at an appropriate level, whether it be technology use or reading comprehension) | | | | |
| Therapist shows that they are promoting participant's independent skills, for example, Can you show me how you would do x? | | | | |
| Therapist does not deviate from the core themes/aims of the session for more than approximately 15% of the session overall | | | | |
| <i>Therapist general skills</i> | | | | |
| Therapist communicates respect to participants in a non-patronizing and sensitive manner | | | | |
| Therapist gives differential feedback | | | | |

APPENDIX B: EXTRACTS FROM THE COMMUNICATE MANUAL

The CommuniCATE Research Project: Enhancing Communication in Aphasia through Technology and Education

Clinician Manual: Using Technology in Reading Therapy

The CommuniCATE project was carried out at City, University of London.

With thanks to our clinical champions, student clinicians, volunteers and participants.

Background

Purpose of this manual

The purpose of this manual is to provide the clinician with strategies and tools to support therapy for people with aphasia using technology for reading. This includes a *clin-*

ician guide and an *aphasia-friendly handbook* on how to therapeutically use technology for reading.

We acknowledge the limitations of this manual relating to software and device upgrades and emphasize that it should serve as a *guidebook not a prescriptive intervention approach*. Most of the manual focuses on the use of Kindles and Claro Read/Speak as that was the software used in our research. The team acknowledges that much of the advice can be applied to other software especially those that use text-to-speech.

The manual is organized into two parts; with the first focusing on how to identify suitable clients and introduce technology for therapy.

The second section focuses on the therapy itself. This section includes examples of goal-setting, session overview and how to adapt reading strategies to technology.

We hope this manual can be used to demonstrate to clinicians the useful features of technology to facilitate reading therapy. We hope it works around some of the barriers that can exist when using technology with clients. And finally we aspire to reduce anxieties around use of technology with clients.

The CommuniCATE project

This resource is an outcome of the CommuniCATE Project run by City, University of London, in conjunction with Barts Health NHS Trust.

The 3-year-long project has investigated the effects of speech and language therapy using computer technologies for people with aphasia, including whether changes in language and communication also improve social participation and quality of life. The project included four strands of therapy focusing on reading, writing, speech production and supported conversation.

The project team have run a number of training sessions for SLTs and volunteers working with the Stroke Association and the feedback from these sessions has been incorporated into the writing of this manual. Our participant user group have helped us to develop aphasia-friendly materials.

The Reading strand aimed to improve access to reading by the use of technology. Treatment combined learning how to operate technologies for assistive reading. This was combined with general reading strategies and encouragement to become more independent with reading and access resources in the community.

Part 1: Reading technology

What is reading technology?

In our project we refer to reading technology as assistive devices that can offer a compensatory approach for reading skills, that is, technology that circumvents the language deficit and utilizes relatively spared language processing skills. This is largely using computers and tablets.

It encompasses a range of technologies, though our project primarily focused on using Kindles, and Claro software on iPad and PC. Other reading technology exists such as other software and platforms, and increasingly more advanced devices are being created such as glasses that can process written text and translate into speech.

With these devices you can read a range of different digital material either from the screen (exactly as you would from paper) or listen to the text (called text-to-speech).

When using a Kindle, you need to have an Amazon account which you can set up here (<https://tinyurl.com/>

[j3qeomp](#)). You can download some books for free, but others cost so you or your client will need to have your or their debit/credit card details to use the books feature. Alternatively, you can use an Amazon gift card if your client has reservations about cyber security.

If using Claro on an iPad you will have to install the app via the Appstore. You will need to create an Apple account using your normal email address. There is a cost attached so you/your client will need to have a card registered or use Apple iTunes vouchers.

We recommend you have separate work accounts for these. This is in line with RCSLT recommendations for professional boundaries using social media.

What are the main features?

The main features of the technology used on the CommuniCATE project involve text to speech, where the device can change written text on the screen into a spoken output. This enables a person with aphasia (PWA) to read using their auditory comprehension to facilitate their reading. This is what makes reading technology advantageous over reading from paper.

There are also additional features that we found helpful for PWA. These include:

Personalizing text presentation: you can alter the size of text, font and line spacing as well as colour of background and text which can make text appear much more aphasia friendly and of preference to the PWA. This alone can often facilitate the PWA to read more/better.

Highlighting and making digital notes: this is akin to 'scribbles in the margin' on paper. The user can highlight words or sections of text and write something about them, for example, paraphrase it to aid comprehension, write reminders to themselves about what it links to and so on. This is useful for those with comprehension and/or memory difficulties.

Using the dictionary: you can see a dictionary definition almost instantly, of every word desired which helps the PWA if they have comprehension difficulties. It removes a lot of the effort that would be required if this were done manually and allows the PWA a more seamless reading experience.

Word Wise: this feature shows a simplified definition of difficult words directly above the word. It can be turned on/off or the sensitivity changed.

X-ray: if available for a book, this gives you the overview of the characters, chapter outlines and key terms. It is very useful for those who have trouble retaining what they have read or recalling who is who in the plot.

Criteria for reading therapy participants on the CommuniCATE project

Twenty PWA took part in the Reading wing of CATE. We formally screened these participants for acceptance onto the project using the Cognitive Linguistic Quick Test (Helm-Estabrooks 2001). We planned to exclude anyone who emerged as severe across the board on the CLQT. None emerged as impaired across the board. Therefore, none was excluded for this reason. It was hypothesized that a severe cognitive impairment as indicated on the CLQT would impair a PWA's ability to learn to use technology.

We carried out a number of formal reading-based pre-therapy assessments for our research, but this is of course entirely at the discretion of the therapist to determine suitability for this type of therapy. This included the Reading and writing subtests from The Comprehensive Aphasia Test (Swinburn et al. 2005), the Reading Comprehension Battery for Aphasia, 2nd Ed. (LaPointe and Horner 1998), the Gray Oral Reading Test, 4th Ed. (Wiederholt and Bryant 2001) and the Reading Confidence and Emotions Rating Scale (Cocks et al. 2013). These were used to help identify the main level of difficulty in reading and results were discussed with the individual.

Other criteria to consider

This is not a prerequisite to using technology or reading therapy, but there are some factors that you may wish to consider:

Individual *impairments* impacting on activities and participation.

Environmental factors.

Technology experience and access.

Relate these to the individual client needs, motivations and *goals*.

We did not conduct in-depth assessment into the type of acquired dyslexia a participant presented with. The therapy worked on a compensatory reading approach and so ultimately a detailed assessment was not required. Some factors that may be worth screening informally include:

Letter matching: matching letters of same case and font.

Letter recognition: distinguishing correct letters from mirror-reversed letters.

Neglect dyslexia: pattern of errors in reading occurring at beginning or end of words.

Attentional dyslexia: errors involving interference/migration of letters from other words, for example, 'flutterfly' for butterfly.

Pure alexia: letter by letter reading: reading letters aloud/very slow reading.

Visual alexia: visual errors, for example, 'better' for butter – see RCBA single word visual distractor subtest.

A typical introductory session plan may look something like this:

| Goal | Activity |
|-------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| For the client to identify what they would like to be reading more of | Use the reading interview and talking mats resource (see Appendices C–E) to facilitate conversation, tick/circle/assign those that may be goals |
| For the client to understand basis of what reading technology is | Use resource in Appendices A and B to help explain to the client what the features of reading technology is, with lots of demonstrations Demonstrate how they work using personally relevant examples, for example, if in the interview the PWA has told you they used to enjoy reading the EastEnders thread on Huffington post, bring this up on a PC and show text-to-speech on Claro Use a Kindle to show how you can transform the presentation of text (e.g., make the font large) |
| For the client to understand differences between devices/software | Use the resource in Appendix F with demonstrations of both a PC/laptop and Kindle |
| For the client to identify how reading technology might fit into what they want to be reading (goals) | From what you have found about what they would like to be reading more of, and what technology they might like, show on a device how these goals could be achieved, for example, show them the bookstore on Kindle to find titles by their favourite authors, demonstrate Claro reading aloud an email and so on |

How to support people with aphasia to use the technology

- We have developed aphasia-friendly guides for Skype alongside this manual. These cover instructions on how to use all features of Skype. When supporting the PWA, use these to facilitate your instructions and provide them with a copy.
- Introduce the tech and software with *lots of demonstrations and relevant examples*, for example, holding interesting things up to the camera, using instant messaging.

- Demonstrate each step *several times* before asking the client to repeat the step.
- Build up *the level of difficulty* dependent on client profile. For example, begin with supporting clients to operate the basic functions of the technology (e.g., turning on, setting up for appropriate camera angle etc.).
- Use *verbal prompts* as well as the aphasia-friendly handbook to support them visually and if necessary, demonstrate the task again.
- Only *move on when the client can perform a function independently* and be prepared to revisit functions taught at each session.
- Once all the basic functions of the technology are mastered *introduce more advanced features if relevant*.
- Use *conversation topics of interest* to the client and relevant to their therapy goals to practice different functions.
- Some clients may benefit from a *diary sheet to record what they have learnt (and/or read)* in sessions or at home. Please see Appendix G for an example.

Part 2: Reading therapy

Therapy programme and activities

We provided a 6-week, 12-session block of therapy to those participants completing the reading strand of the CommuNICATE project. The participants also received up to two technology training sessions each lasting 1 h.

The therapy focus tended to change across the block, although a typical therapy session would aim to include activities based on the three goal areas of technology, participation and communication. We have described some brief therapy ideas in the goal-setting examples above. Here we suggest what we found to be the typical programme for a therapy block, with some further suggestions for therapy activities.

Technology training and sessions 1–3: largely involve setting up the technology, support for learning basic skills and setting broad goals (e.g., what do they want to be able to do at the end of this block?). Trialling different features of technology to see what is beneficial. How much do the different features help them (e.g., do they benefit from having lines spaced further apart?). What is their capacity for learning to use the technology independently and how much support will they need (e.g., with the visual guides can they practise at home? Are they motivated to do so?). Goals might need to be negotiated and modified a lot during the early sessions.

Activities: setting up an Amazon account if necessary, practical work for technology tuition (assess the best way they learn these skills—some like to watch and listen, others need to be ‘hands on’), introducing the aphasia-friendly

guides, lots of demonstrations of basic features. Some may need basic technology training (e.g., turning on a device, using a touch screen) so always ensure you pitch activities at appropriate level and pace.

Sessions 4–9: the main intervention period, which focused on specific activities to achieve their goals, with ongoing technology support (i.e., continually extending skills if appropriate). This will vary massively depending on their goals, but it is good if you now have an idea of how the different features can be tailored to meet goals.

Sessions 10–12: largely facilitating the PWA to maintain new skills after their therapy ends. This may be across all goal areas, for example, how will they purchase new books on the Kindle (technology goal), do they have the skills, knowledge and confidence to regularly hold conversations about the news (participation goal), and have they used text-to-speech to read material other than that identified in therapy sessions? (Communication goal).

Adapting reading strategies in assistive technology

There are many strategies that SLTs use to make reading more accessible for PWA. This therapy was largely based on learning to use a piece of technology and exploring the support offered by the technology. Below we outline some of these strategies. These can be incorporated into your session activities.

| Reading therapy strategies typically used in ‘paper’ reading work by an SLT | Translation of functional reading strategies to using features of technology for the same purpose |
|-----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Large and clear writing | Change the text font, size and line spacing |
| Non-distracting background | Adjust the screen to ‘text only’. Use of ‘reading view’ from webpages to simplify the layout |
| Reading aloud | Text-to-speech |
| Interim summaries/reviews | X-ray on Kindle Notes made on Kindle Writing on Claro |
| Use of images | Use Google images/links on articles. Integrating the technologies with other apps |
| Finding definitions in dictionary | Automatic dictionary ‘look up’ features. Use of word wise |