Originalbeitrag

Investigating treatment fidelity in a conversation-based aphasia therapy

Heilemann, Claudia [1]; Best, Wendy [1]; Johnson, Fiona [1 2]; Beckley, Firle [1]; Edwards, Susan [1 3]; Maxim, Jane [1]; Beeke, Suzanne [1]

EN | Abstract

When conducting a speech and language therapy intervention study, one essential focus is the evaluation of outcomes. The therapy itself and its actual delivery, often discussed in the context of treatment fidelity (TF), are aspects which are mostly carried out in the background. According to Cherney et al. (2013), however, TF is an important component of any behavioural treatment study and should therefore be investigated.

This article presents a study of TF embedded in a wider research project that evaluates a new conversation-based therapy for people with agrammatic aphasia and their conversation partners (Beckley et al., 2013; Beeke et al., 2011; Beeke et al., 2014). The therapy is called Better Conversations with Aphasia (BCA). Using the concept of TF, the degree to which BCA was delivered as planned can be measured. A pilot version of a BCA-specific observational fidelity tool was developed, based on a conceptual model of TF (Carroll et al., 2007), on practices reported in the TF literature and on the generic therapy session plans. The results indicate that, in terms of adherence to the therapy content, a high fidelity level (91.9 %) was reached for BCA. This article will also report on the degree to which the therapist showed desired behaviour associated with the delivery of BCA, and findings from an inter-rater reliability investigation of the fidelity tool. It concludes with reflections on the importance and value of TF investigations in speech and language therapy intervention.

[1] Division of Psychology and Language Sciences, University College London, London, UK
[2] Guys & St Thomas' NHS Foundation Trust, London, UK.
[3] School of Psychology and Clinical Language Sciences, University of Reading, UK.

DE | Zusammenfassung

Die Evaluierung einer Therapiestudie besteht typischerweise aus der Messung des tatsächlichen Therapieeffekts. Die Frage, inwieweit hierbei die ursprünglich vorgesehene Therapie angewendet wurde (z.B. wie in einem Therapiemanual beschrieben), wird hingegen oft vernachlässigt. In der englischsprachigen Literatur wird in diesem Zusammenhang häufig von dem Konzept der treatment fidelity (TF) gesprochen. Das Konzept der TF wird nach Cherney et al. (2013) als eine wichtige Komponente einer verhaltenstherapeutisch orientierten (Sprach-)Therapiestudie angesehen.

In diesem Artikel wird eine Studie vorgestellt, die in ein übergreifendes Forschungsprojekt (Beckley et al., 2013; Beeke et al., 2011; Beeke et al., 2014) eingebettet ist. Das Konzept der TF wird auf eine konversationsorientierte Therapie (Better Converations with Aphasia; BCA) für Menschen mit Agrammatismus und deren Konversationspartner angewendet. Ziel ist es, den Grad der Konformität mit der ursprünglich vorgesehenen Therapie zu ermitteln. Hierzu wurde eine Pilotversion eines BCA-spezifischen Beobachtungsinstruments (fidelity tool) entwickelt, basierend auf einem speziellen TF-Modell (Carroll et al., 2007), weiteren Methoden aus der TF-Literatur sowie BCA-spezifischen Therapieplänen. Die Analyse der Videoaufnahmen von Therapiesitzungen mithilfe des Beobachtungsinstruments zeigt einen fidelity score von 91,9 %, was laut Literatur auf eine hohe Therapiekonformität hindeutet. Zudem werden Ergebnisse zu der Qualität der Therapieübermittlung vorgestellt sowie die Inter-rater Reliabilität des fidelity tools diskutiert. Insgesamt demonstriert diese Studie die Wichtigkeit und das Potential einer fidelity Evaluation anhand der Anwendung des TF Konzepts auf die BCA Therapie.

1. Conversation-based therapy approaches

Aphasia is traditionally defined as an acquired language impairment resulting from a brain lesion in the language dominant hemisphere (Papathanasiou, Coppens, & Potagas, 2013). Generally, aphasia can be divided into several subtypes characterised by typical symptom complexes (Huber, Poeck, & Springer, 2006). One of these symptom complexes consists of the use of simplified syntactic structures, omission of morphology and few if any verbs. Characterised by non-fluent speech production, this symptom complex is known as agrammatism. It will be the focus of this article.

Beeke et al. (2011) suggest that targeting the grammar of everyday conversations may be more effective in improving the spontaneous speech of people with agrammatism than targeting decontextualised language (e.g. the language used when describing pictures). Beeke, Wilkinson and Maxim (2007) demonstrated that the language

Partner Scheme› (McVicker, Parr, Pound, & Duchan, 2009), developed at Connect, the communication disability network, in London, to train volunteers to converse with PWA who experience social isolation (see also Swinburn, McVicker, & Pearce-Willis, 2011). Another is ‹Supported Conversation for adults with Aphasia› (SCA; Kagan, 1998a, 1998b; Kagan & Gailey, 1993), which also focuses on conversation partner training. Table 1 shows an overview of selected conversation-based therapy approaches for aphasia.

It should be noted that the examples included in Table 1 do not exclusively target agrammatic speech and that they are not intended to be a complete enumeration of all existing conversation-based therapy approaches. For the most part, studies using these programmes focus on training the CP rather than the PWA (for a systematic review, see Simmons-Mackie, Raymer, Armstrong, Holland, & Cherney, 2010).

Table 1: Conversation-based therapy approaches for aphasia (chronologically ordered)

Therapy approach	Abbreviation	Authors (Years)
‹Conversation Coaching›	-	Holland (1991, 1997)
‹Communication Partners›	-	Lyon et al. (1997)
‹Supported Conversation for adults with aphasia›	SCA	Kagan (1998a, 1998b), Kagan & Gailey (1993)
‹Supporting Partners of People with Aphasia in Relationships and Conversation›	SPPARC	Lock et al. (2001a)
‹Conversation Partner Scheme›	-	McVicker et al. (2009)
‹Partners van Afasiepatienten Conversatie Training› (Dutch adaptation of SPPARC)	PACT	Wielaert & Wilkinson (2012)
Swedish adaptation of SPPARC (conversation partner training)	-	Saldert, Backman, & Hartelius (2013)

output produced by speakers with agrammatism in everyday conversations is different to speech elicited by a task-based language assessment, and they therefore suggest that focusing therapy on elicited speech may be limiting generalisation.

Since the 1990s, researchers have developed and tested various approaches that target aphasic conversation directly. So-called conversation-based (sometimes referred to as interaction-focused or pragmatic-oriented) therapy approaches focus on the conversation behaviour of conversation partners (CPs) of people with aphasia (PWA), or the CP and the PWA as a couple, with the objective of enhancing daily, natural conversations, by practising the use of conversation strategies (for a review, see Wilkinson & Wielaert, 2012). One example is the ‹Conversation

The Better Conversations with Aphasia therapy programme

Beeke and colleagues (Beckley et al., 2013; Beeke et al., 2011; Beeke et al., 2014) have designed a conversation-based therapy approach, called BCA, based on SPPARC (Lock et al., 2001a). SPPARC is a conversation training that applies a qualitative research method, called Conversation Analysis. It consists of three progressive main steps: (1) to raise awareness of the broad idea of conversation and conversation behaviour (such as ‹the aim of a turn› or ‹overlapping talk›) in relation to aphasia, (2) to raise awareness of a CP’s own conversation behaviour, and (3) to facilitate the identification and use of strategies for change (e.g. waiting until the PWA’s turn is finished before speaking). The key techniques included in

the therapy (video feedback, role-play and conversation activities) are based on a model of experiential learning (Kolb, 1984; see Beckley et al., 2013 for an overview). Several studies found mostly qualitative indicators for SPPARC's efficacy (Burch, Wilkinson, & Lock, 2002; Lock et al., 2001a, 2001b; Wilkinson, Bryan, Lock, & Sage, 2010) and showed that changes in a CP's behaviour could lead to indirect changes in a PWA's conversation. SPPARC is designed to be delivered either in a group setting (that is, a group of CPs) or with the PWA and the CP together as a couple; either way, it focuses on directly changing the CP's conversation behaviour.

Table 2: Session goals of the BCA therapy programme

Session #	The overall goal(s) of the session is/are...
1	...to raise the dyad's overall awareness of conversation.
2	...to raise the dyad's awareness of different aims of turns.
3	...to raise the dyad's awareness of repair in general AND to help the dyad to identify their own patterns of repair.
4	...to identify patterns of turn building in the PWA's own conversation AND for the PWA to select strategies for change and to experience them within a structured task.
5	...to identify patterns of turn building in the CP's own conversation AND for the CP to select and practice strategies for change.
6	...to facilitate the identification and implementation of strategies for change in relation to topic.
7	...to facilitate the implementation of strategies for change.
8	...to support the dyad to implement strategies for change.

BCA uses Kolb's adult learning model (1984) as its basis, as does SPPARC. However, Beeke and colleagues created two significant changes from SPPARC in their BCA therapy programme. First, they focus on conversational difficulties resulting from agrammatism instead of aphasia in general (Beeke et al., 2011). Secondly, in BCA, the PWA is an active participant throughout all sessions. Instead of solely targeting the CP (as is common in most of the previous research in this area), BCA is designed to train a PWA as well as their CP to use conversation strategies. This was done in order to discover whether a PWA can learn from conversation therapy (see Beckley et al., 2013). This adaptation led to the creation of a therapy session focused solely on strategies for the PWA (e.g. use of gesture, key words, writing and drawing in a conversati-

onal turn), plus linked therapy handouts and activities. BCA is designed as a programme of 8 weekly sessions of around 1.5 hours in length taking place at the clients' home (Beckley et al., 2013). The overall aim of BCA is to educate the participants about the effects of agrammatism on conversation, and teach conversation strategies to allow a PWA to produce more successful turns, which should increase mutual understanding between a dyad (i.e. a PWA and their frequent CP). The specific goals for each BCA therapy session are illustrated in Table 2. The therapy programme itself is freely available as part of an e-learning package on UCLeXtend (https://extend.ucl.ac.uk). The change in conversations associated with the therapy is expected to be both quantitative and qualitative in nature (Beeke et al., 2011).

In summary, conversation-based therapy programmes such as CP training have become popular clinically. Yet they can be regarded as complex interventions (Beeke et al., 2014; Cherney et al., 2013), with interacting components and individually variable outcomes. Thus, there is a need to explore the delivery of therapy (see also Medical Research Council-guidelines: Craig et al., 2008), especially with regard to the research quality of an intervention study. Cherney et al., 2013, in their paper about the methodological quality of studies on CP training in aphasia, speak of treatment fidelity (TF) as an important element of a behavioural treatment study. This concept will be explored next.

2. The concept of treatment fidelity (TF)

When building an evidence base for a new and complex behavioural treatment – like BCA – researchers need to report on the treatment itself (Craig et al., 2008) as well as the therapy outcomes. In the context of designing, implementing and evaluating treatments, one key element is TF (Hennessey & Rumrill, 2003), a measure of the reliability of the provision of a treatment (Hinckley & Douglas, 2013). Literature on the concept of TF has been published in different research areas, such as psychology, education and medicine. Particularly in the last 20 to 30 years many researchers have developed fidelity assessments (e.g. Bellg et al., 2004; Carroll et al., 2007; Waltz, Addis, Koerner, & Jacobson, 1993; Yeaton & Sechrest, 1981). The literature discussed in this article is predominantly from studies that investigate aspects of TF from the field of psychology, where the concept originated (Di Rezze, Law, Gorter, Eva, & Pollock, 2012), because this field has most in common with conversation-based speech and language therapy. Compared to this rich body of research, little literature on TF exists in the area of speech and language therapy studies. This may

be because TF issues are usually reported in the context of higher-level study designs such as randomised control trials (see e.g. Godfrey, Chalder, Ridsdale, Seed, & Ogden, 2007). Such trials are difficult to design and execute in speech and language therapy and are therefore relatively rare (Cherney et al., 2013). Some papers, however, do report on some aspects of TF (e.g. monitoring and documenting fidelity) especially when assessing complex speech and language therapy interventions (see e.g. Adams, Lockton, Gaile, Earl, & Freed, 2012). Last year, Hinckley & Douglas (2013) published the first review of the importance of TF and the frequency with which it is reported in studies related to aphasia treatment. This underlines the growing attention to TF in the field.

2.1 Definition

TF is a term that encapsulates a concept originally known as treatment integrity. It assesses whether a treatment or therapy was delivered as intended (Hennessey & Rumrill, 2003). Other terms to describe TF are procedural reliability, intervention fidelity, implementation fidelity, program-(me) fidelity, treatment adherence process research and therapist's or clinician's adherence or competence. This range of terminology reflects variability in definitions of TF and leads to confusion in terms of what exactly should be measured when it is assessed (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012). In this article, the term ‹TF› will be used consistently.

The specific term treatment fidelity was first employed by Moncher and Prinz (1991). Lichstein, Riedel, and Grieve (1994) added two aspects to previous definitions: receipt of treatment (i.e. that the client understands and uses the skills that are delivered) and enactment (i.e. that the client implements these skills into daily life). This addition of client-focused behaviour enlarges the traditional understanding of treatment integrity related to adherence (Were all components of the therapy delivered?) and competence (In what way has the therapy been delivered?).

2.2 Measurement

Methods of measuring TF vary according to intervention type (see e.g. Chan et al., 2004; Kiran & Thompson, 2003; Lichstein et al., 1994). In general, TF measurement can take place either once (e.g. at an initial stage of a new therapy or at any point after implementation), or repeatedly (Bond, Evans, Salyers, Williams, & Kim, 2000).

Fidelity measures also vary in their aims. For example, researchers might aim to document adherence, or to discriminate treatments which share similarities, to synthesise a body of research, or to identify core ingredients of a therapy (e.g. specific treatment targets or therapeutic techniques).

Common fidelity tools include direct observation, observer and self-report check-lists, indirect observation via video and ratings by experts based on documentation data (Kaderavek & Justice, 2010; Mowbray et al., 2003; Resnick et al., 2005). Fidelity tools are easier to develop if «detailed practice manuals» for a therapy exist (Bond et al., 2000: p. 78).

In order to investigate TF as precisely as possible, many researchers use a multidimensional approach to fidelity, including structural elements (adherence to the therapy programme) and procedural aspects (quality of the delivery, emotional climate in a session, therapy principles) (e.g. Hasson, Blomberg, & Dunér, 2012; Odom et al., 2010).

The study presented in this article is underpinned by benchmarks in relation to fidelity procedures, synthesised from the TF literature (see Table 3).

2.3 Conceptual model of TF

The framework created by Carroll et al. (2007) is useful for integrating TF into the research process and describing the ingredients and potential influencing aspects of TF (Hasson, 2010). This so-called Implementation Fidelity Framework (IFF), presented in Figure 1, includes traditional components of TF (e.g. adherence to therapy content) as well as so-called moderating factors, which are expec-

Table 3: Summary of suggested benchmarks in relation to fidelity procedures (based on a synthesis of TF literature)

Question regarding fidelity procedures	Suggestion derived from the TF literature
Amount of therapy sessions to be checked?	15% - 40%, randomly selected
Type of rating scale used in order to conduct a fidelity check?	Likert-type scale or occurrence/non-occurrence
Characteristics of rater(s)?	Trained, independent, familiar with the intervention
‹High› fidelity level?	80% and above
Amount of sessions to be checked by a second rater?	10-30%
Acceptable level for inter-rater percentage agreement?	70% and above
Computing inter-rater reliability (IRR) of a fidelity tool?	Intra-class correlation, Kappa

ted to influence the degree of overall TF. It served as a foundation for the present study. The IFF components will now be outlined.

According to the IFF, the core of TF is adherence, with subcategories content, coverage, frequency and duration. The assessment of adherence involves ascertaining whether the main elements of an intervention have been implemented as intended and if the participants received these elements as often and for as long as designed by the developers of a therapy. Adherence may be affected by moderating factors. Potential moderating factors include the complexity of the intervention, facilitation strategies such as training of therapists or preparation of a manual, the quality of therapy delivery and participant responsiveness.

The link between TF and therapy outcomes is visualised with the help of a broken line (see Figure 1). This indicates that the relationship between a treatment and its outcomes is external to TF, although, according to Carroll et al. (2007), the degree of TF can in theory affect the outcomes.

The notation «identification of ‹essential› components» (see Figure 1) implies that the aspect of programme differentiation in the model, other than suggested in the literature (e.g. Moncher & Prinz, 1991), is regarded as separate from TF.
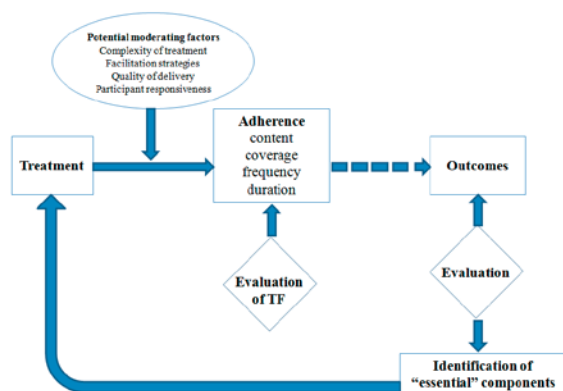

Figure 1: IFF (adapted from Carroll et al., 2007: p.4)

3.   Research questions

As outlined above, the evaluation of TF is an important part of the methodological quality of a study targeting behaviour change. With regard to future investigations of conversation-based therapy approaches, it is therefore necessary to develop adequate tools to measure TF.

A multidimensional fidelity tool was designed, based on the first author's review of the TF literature, with the intention of capturing as complete a representation of TF as possible for the research project evaluating BCA. The

IFF (Carroll et al., 2007) served as a conceptual model in developing the following research questions:

1. To what degree were the planned components of the BCA therapy programme delivered to the participants of the main research project? [adherence]
2. To what degree did the behaviour of the therapist reflect desired BCA intervention principles? [therapy delivery]

4.   Method

The data for this study stem from a research project (Beeke et al., 2011) executed at University College London (UCL) in which the effectiveness of the BCA therapy programme was tested. The study design of that project comprised a case series evaluation (for details see Beckley et al., 2013).

4.1   Participants

Participants were recruited for the main project from speech and language therapists working in the National Health Service, aphasia support groups, private speech and language therapists and university aphasia clinics. Seven dyads from the main project, in each case one person classified as having agrammatic aphasia and his or her significant other (spouse or family member), were analysed in the present study. All people with aphasia (PWA) had had a left hemisphere stroke at least 6 months prior to involvement in the project. They were classified as agrammatic according to their spoken output in conversation with the research speech and language therapist (SLT) and from their verbal description of the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (BDAE, Goodglass & Kaplan, 2001). All participants were English native speakers. Each dyad was treated by the same research SLT, who had more than 10 years of clinical experience. PWA consisted of 3 females and 4 males (time since stroke Ø 40 ± 16 months, age Ø 56 ± 10 years, age left education Ø 18 ± 2 years) and their 7 CPs (5 females, 2 males).

4.2   Description of the data

During the main research project, each therapy session was videotaped using a Panasonic digital video camera. The dyad and the SLT are clearly visible on screen. Data analysed here consist of a randomly selected sample of 25 % of all videotaped therapy sessions (N=14; two sessions out of 8 for each of 7 dyads), equating to a total of 17.3 hours of data. In addition, written notes by the research SLT were consulted for information on the actual content of individual sessions. Data were collated retrospectively allowing TF to be studied independently.

The generic BCA session plans were used to construct a fidelity tool since a BCA manual has yet to be developed.

4.3 Fidelity tool

A pilot fidelity tool was devised, covering the IFF (Carroll et al., 2007) components of adherence, with subcomponents content and dose, and potential moderating factors. As the essential components of BCA, i.e. those that are expected to create therapeutic change, are not known yet, fidelity to each component of the therapy programme, was examined (Carroll et al., 2007). The tool consisted of a ‹procedural› section, created to carry out the main fidelity check, a ‹qualitative› section, to analyse therapy delivery, and a ‹client-focused› section, to capture participant responsiveness (i.e. opinions of the clients expressed during a therapy session). The procedural and qualitative sections will now be described in more detail, because the present article focuses on adherence to therapy content and therapy delivery. Therefore, the client-focused section stays in the background of the present article.

4.3.1 Procedural section

The items included in the procedural section of the tool were based on the a priori constructed generic session plans that exist for each session, i.e. they were specific to the activities and materials of BCA therapy (e.g. ‹The therapist showed a video example of a successful conversation repair›). They were therapist-oriented (following the recommendations by Hogue et al., 1996 to rate only therapist behaviour when assessing adherence). The section was designed to enable an external rater to fill it in on observation of therapy videos. Almost all of the items identified adherence, but some items were expected to indicate both adherence and quality of delivery (e.g. when the therapist leads a discussion with the dyad). The three-point Likert-type rating scale that was applied to the procedural section distinguished between therapy content that was fully, partly or not delivered as planned (1=fully delivered; 0.5=partly delivered; 0=not delivered). This permitted a rater to record items that were not fully delivered (e.g. because of a certain participant reaction), and so the interactive nature of the therapy was captured.

4.3.2 Qualitative section

With regard to the present study, it is important to know which therapist skills are associated with the delivery of BCA. A list of principles reflecting desired therapist behaviour was created from the SPPARC manual and the adaptations made for BCA. Thus, ideally, the therapist:

- supplies individualised advice (based on analyses of conversation between the PWA and the CP)

- guides the dyad to make their own choices
- focuses both on the PWA and on the CP, so that the PWA and the CP have equal roles during the session
- avoids making judgments about what conversation patterns the dyad should retain or change
- uses active listening skills
- gives skilful summaries of what has been said
- expresses warmth and empathy towards the dyad
- affirms and encourages the dyad

These fundamental principles were listed for rating in the qualitative section of the tool. Following Chan et al. (2004), the rating scale used distinguishes between therapist behaviour which the rater observes not at all (0), occasionally (0.5) or most of the time (1).

Another aspect that was included in the qualitative part of the fidelity tool was whether the overall aim of the session was judged to be met. A statement for each session (e.g. session 2: ‹The overall aim of the session, to raise the dyad's awareness of different aims of turns, was achieved›) was listed in order to be rated with 0 (I don't agree), 0.5 (I partly agree) or 1 (I fully agree).

4.4 Procedures

A mixed-methods approach was conducted in order to answer the research questions. Guided by the IFF (Carroll et al., 2007), data concerning adherence and potential moderating factors were obtained via observation of the video-recorded therapy sessions by the first author of the present article. Additionally, secondary data sources (an email survey and document analyses such as notes on individual session plans or dates on video tapes) were conducted in order to examine aspects of TF. A second rater (a speech and language therapy graduate of UCL already familiar with the main research project) observed and rated 20 % of the sessions coded by the first author (N=3) in order to investigate inter-rater reliability (IRR) of the procedural and the qualitative sections of the fidelity tool. The aim was to uncover whether the fidelity tool can be reliably implemented by different independent raters in the same way using the same video samples.

5. Results

The IRR analysis revealed that the overall percentage agreement of the procedural and qualitative section was similar (86.8 % and 87.5 % respectively), but higher variability was evident in the qualitative section of the fidelity tool. More detailed results regarding IRR are summarised in Appendix 1.

5.1 Adherence

‹To what degree were the planned components of the BCA therapy programme delivered to the participants of the main research project?›

Here the focus is what Carroll et al. (2007) in their model refer to as ‹content›. Across the sample of 14 sessions, the overall number of observations (i.e. items / therapy activities) was 232. Numbers of planned activities (e.g. talking through a specific handout) varied across each of the sessions. It was not possible to carry out the fidelity check on session N° 8 for dyad N° 3, because there were technical problems with the video recording, and the handwritten session notes by the SLT did not include enough information on the missing items to complete the ratings. The data for this session were therefore removed from further analysis.

The results of the ratings for the procedural section of the fidelity tool are shown in Table 4. The mean number of observations per dyad was 32 (SD=9.7) with a range between 18 for dyad N° 3 (session 5) and 41 for dyad N° 1 (sessions 4 and 6). Per session across all dyads, the mean number of observations was 18 (SD=5.2; range: 5-25).

The final data include a total of 227 observations over 13 sessions which would allow a maximum overall fidelity score of 227. The achieved score given by the rater is 208. This represents 91.9 % (SD=3.9), i.e. the overall fidelity score across all dyads related to the IFF subcategory ‹con-

tent› is 91.9 %, based on an analysis of 23 % of the 56 therapy sessions.

A closer inspection of the dyad-specific fidelity scores reveals four of these are above 90 % (for dyads N° 1, 2, 3 and 6). For two dyads, the scores lie just below 90 % (dyads N° 4 and 5). The fidelity score for dyad N° 7 is the lowest at 86.1 %. It must be kept in mind that the scores of the dyads are based on different numbers of observations, as these varied between sessions.

Averaged across sessions and dyads, 88.5 % of the items were given a rating of 1 (fully delivered, N=201), 6.2 % of the ratings indicated partial adherence (corresponding to a rating of 0.5, N=14) and 5.3 % of the items were given a rating of 0 (not delivered, N=12).

5.2 Therapy delivery

‹To what degree did the behaviour of the therapist reflect desired BCA intervention principles?›

The results for the qualitative section of the fidelity tool are reported in Table 5. Here the focus is what Carroll et al. (2007) refer to as ‹quality of delivery›.

Again, session N° 8 for dyad N° 3 was removed from the analysis (see section 5.1). The results of the TF check for the qualitative section, based on the remaining 13 sessions, show scores near ceiling: They range between 90 % (dyad N° 7) and 100 % (dyads N° 3, 4 and 5). Overall, the degree to which desired therapist behaviour was present was 96.7 % (SD=4.1).

**Table 4:** Dyad-specific fidelity scores and overall fidelity score

Dyad #	Session # checked	Activities planned, Maximum score	Activities delivered, Actual score	Percentage $\frac{Actual\ score}{Maximum\ score} *100$
1	4 6	41	37.5	**91.5**
2	2 3	38	35.5	**93.4**
3	5	18	17.5	**97.2**
4	1 6	39	35	**89.7**
5	7 8	19	17	**89.5**
6	3 4	36	34.5	**95.8**
7	3 4	36	31	**86.1**
Total		**227**	**208**	
Mean		**32.4**	**29.7**	**91.9**
Standard deviation	9.7	8.7	3.9	

Table 5: Dyad-specific quality scores and overall quality score

Dyad #	Session # checked	Maximum score of desired behaviour	Actual score	Percentage $\frac{Actual\ score}{Maximum\ score} *100$
1	4 6	19	17.5	**92.1**
2	2 3	19	18.5	**97.4**
3	5	10	10	**100**
4	1 6	16	16	**100**
5	7 8	16	16	**100**
6	3 4	20	19.5	**97.5**
7	3 4	20	18	**90.0**
Mean		**17.1**	**16.5**	**96.7**
Standard deviation	3.6	3.1	4.1	

The frequency with which each item was rated with 0 (not at all), 0.5 (occasionally) or 1 (most of the time) can be found in Figure 2. The item «The therapist avoided judgements about what conversation patterns the dyad should retain or change» was most frequently rated at 0.5 compared to other items. No single item in the fidelity check, however, was rated at 0. Thus, for this section of the TF tool, Likert-scaling turned into a binary rating system. The most consistently rated qualitative items were «The therapist used active listening skills», «The therapist gave skilful summaries of what has been said», «The therapist expressed warmth and empathy towards the dyad» and «The therapist affirmed and encouraged the dyad», each of which reached the maximum score of 13 across all observations.
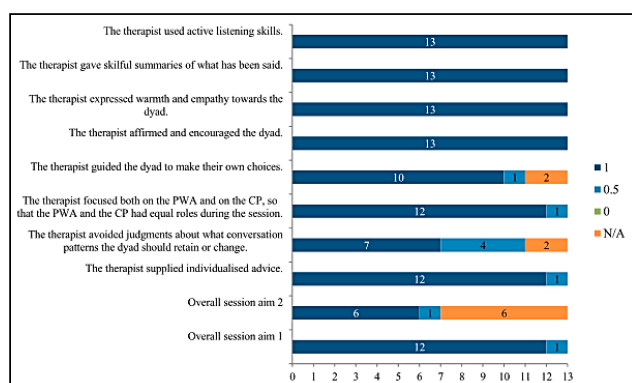
Figure 2: Frequency of ratings (0=not at all; 0.5=occasionally; 1=most of the time) of therapist behaviour across the observed sessions (N=13); N/A=not applicable

6. Discussion

6.1 Adherence

Across the observed therapy sessions, fidelity to the session plans (score: 91.9 %) can be regarded as high, according to the published TF literature. This indicates that BCA therapy delivery was consistent with the prototype therapy as designed by the main project team. It is therefore likely that the so-called active ingredients of BCA therapy have been delivered with high fidelity, although further research is necessary in order to identify these.

Averaged across all observed sessions, the percentages of items that were given a rating of 1 (88.5 %), 0.5 (6.2 %) and 0 (5.3 %) are similar to or better than the results reported in previous literature. For example, Lewinsohn et al. (1990) stated that 78 % of items were given a rating of 2 (corresponding to a rating of 1 in the present study), 17 % a rating of 1 (corresponding to 0.5) and 5 % of the ratings indicated no compliance. However, it is important to keep in mind that the design of the study by Lewinsohn et al. (ibid.), the nature of the intervention and the small number of items in the fidelity tool (11) differ from the current investigation.

Dyad N° 7 showed a lower individual fidelity score (86.1%) compared to the other six dyads (89.5 % and above). This score can be explained by the relatively high proportion of 0 and 0.5 ratings given to items observed. This shows that even when only one experienced therapist is delivering an intervention (which can be described as an «ideal situation», according to Hennessey & Rumrill, 2003:p. 124), there appears to be a certain amount of variety in therapy delivery. Moreover, it shows that a fidelity evaluation can serve as a tool to uncover certain participants for which therapy was delivered in a different way compared to the majority. It also points to the importance of measuring fidelity when there is more than one therapist delivering an intervention. In this regard, it is important to provide appropriate training to all therapists and use strategies such as supervision or a detailed manual in order to ensure that each therapist delivers the planned therapy.

One issue worth noting in this study is the selected sample. Interpretations have to be handled with caution because it could be argued that the final sample (23 % of all sessions) is not representative of the whole therapeutic process, although guidelines from the TF literature concerning the amount of therapy sessions to check were followed. Also, each dyad-specific fidelity score is based on a different number of observations, depending on the individual sessions checked.

The BCA therapy can be described as highly interactive, which is why some constituents of the procedural section of the fidelity tool are worded in a rather open way (e.g. «The therapist had a discussion with the dyad on how aphasia affects conversations»). The data reveal that a relatively high percentage of the 0.5-ratings given by rater 1 (corresponding to ‹partly delivered›) consisted of items belonging to the major domain of having a discussion. This could be a sign of problems for the rater with the definition of these items (the issue of specificity of treatment components is also discussed in Whyte & Hart, 2003). This issue could be further examined in the future, for example, by assessing the degree of complexity of the BCA therapy (e.g. with the help of a survey for experts in the specific therapy approach, see also Carroll et al., 2007) or by rewording or concretising these items.

A general issue in the context of the procedural section of the fidelity tool is the influence which each of the procedural items has on the overall fidelity score. As each single element from the generic session plans was included, there is no variation in their weight related to the overall fidelity score (i.e. each item is equally weighted). By including an estimate of the therapeutic potency of each item, hypothesised active ingredients could be in-

vestigated for their potential to influence fidelity more than elements which might be regarded as less important. In general, future research should aim to develop a shorter fidelity tool reflecting the most important aspects of the BCA therapy approach.

6.2 Therapy delivery

Items in the qualitative section of the fidelity tool are an attempt to describe desired therapist behaviour associated with the delivery of the BCA therapy. However, it is important to acknowledge the complexity of this aspect of TF. High quality scores (ranging from 90 % to 100 %) were found for the seven dyads using a 3-point Likert scale. This ceiling effect could theoretically reflect rater bias. Consequently, it could be advantageous to use a 5-point Likert scale. This may lead to a higher degree of differentiation in the rating of desired therapist behaviour. Another possibility would be to create clear definitions and examples that correspond to the 0, 0.5 or 1-ratings for each qualitative item, to clarify how to use the scale. Furthermore, the challenge for future investigations is to consider whether there are additional behaviours (not currently listed in the TF tool) that are key to competent BCA therapy delivery. One possible skill relates to the ability to select appropriate positive and negative samples of conversation to show a dyad during therapy, as this underpins a vital component of the intervention - the identification of facilitator and barrier conversation behaviours to target in therapy. As yet unidentified aspects of therapy delivery such as this might also have an impact on future studies where more than one therapist is delivering BCA therapy. This highlights the need for comprehensive SLT training when preparing for a reliable delivery of the BCA therapy. One resource that begins to answer such a need can be found at https://extend.ucl.ac.uk, a free e-learning resource based around BCA to help SLTs to plan, carry out and evaluate conversation therapy.

The qualitative item which was most frequently rated with a score of 0.5 was «The therapist avoided making judgments about what conversation patterns the dyad should retain or change». This suggests that the therapist showed the desired behaviour only ‹occasionally›. On the other hand, the coding of this item might reflect the suitability of the clients for the intervention, since it is important in the therapeutic exchange for the therapist to react to client behaviour. For example, if a client is not able to choose what conversation strategy to work on, as a consequence the therapist is more likely to have to judge herself which patterns or strategies should be changed or retained.

7. Conclusion

The fidelity and therapy delivery levels for BCA therapy reported here can be regarded as high, whilst at the same time reflecting a certain amount of flexibility in therapy delivery, which is likely to be desirable for such an interactive therapy approach. These findings contribute to the group outcomes of the main research project by providing evidence that each of the seven dyads received the BCA therapy as originally planned. This study underlines the growing prominence of TF in the field of speech and language therapy, especially with regard to ensuring the methodological quality of research reports. However, valid fidelity tools need to be created for speech and language interventions, of which conversation-based therapy approaches represent one example. In conclusion, this study has demonstrated the multifaceted nature of TF, and its importance and value for the evaluation of a complex speech and language intervention for aphasia.

Appendix 1

Inter-rater reliability of the independent ratings for the procedural and the qualitative section of the fidelity tool averaged across the three sessions observed by both raters

Section of the fidelity tool	Percentage agreement	Range	ICC
Procedural section (adherence/content)	86.8%	80.0-92.9%	.674 (good)
Qualitative section (therapy delivery)	87.5%	62.5-100%	.258 (poor)[a]

[a] This ICC value is not an average but reflects the ICC of one session, session N° 4 for dyad N° 7 (there was a lack of variability in the raw data of the other two sessions so that ICC could not be calculated for all of the three sessions).

IRR was examined for the pairs of dual-rated video-recorded therapy sessions. Both percentage agreement and an intra-class correlation (ICC; two-way random, consistency, single-measures; Shrout & Fleiss, 1979) were performed, the latter to correct for chance agreement (Hallgren, 2012). The table below summarises the results of the IRR investigation. The overall percentage agreement of both sections of the fidelity tool is similar, but the range indicates higher variability in the qualitative section. Averaged across rated sessions, ICC values indicate good agreement across raters for the procedural section, but poor agreement for the qualitative section. The latter could be due to the fact that percentage agreement for the specific session which built the foundation for this ICC value of .258 was only 62.5 %.

Contact | Heilemann, Claudia; UCL Division of Psychology and Language Sciences Chandler House 2 Wakefield Street London WC1N 1PF UK; c.heilemann@ucl.ac.uk

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 Communication Disorders, 47 (3), 245-256.

Beckley, F., Best, W., Johnson, F., Edwards, S., Maxim, J., & Beeke, S. (2013). Conversation therapy for agrammatism: exploring the therapeutic process of engagement and learning by a person with aphasia. International Journal of Language and Communication Disorders, 48 (2), 220-239.

Beeke, S., Wilkinson, R., & Maxim, J. (2007). Grammar without sentence structure: A conversation analytic investigation of agrammatism. Aphasiology, 21 (3/4), 256-282.

Beeke, S., Maxim, J., Best, W., & Cooper, F. (2011). Redesigning therapy for agrammatism: initial findings from a conversation-based approach. Journal of Neurolinguistics, 2, 222- 236.

Beeke, S., Beckley, F., Johnson, F., Heilemann, C., Edwards, S., Maxim, J., & Best, W. (2014). Conversation focused aphasia therapy: investigating the adoption of strategies by people with agrammatism. Aphasiology. DOI: 10.1080/02687038.2014.881459.

Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M., Ogedegbe, G., et al. (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.

Bond, G.R., Evans, L., Salyers, M.P., Williams, J., & Kim, H.-W. (2000). Measurement of Fidelity in Psychiatric Rehabilitation. Mental Health Services Research, 2 (2), 75-87.

Burch, K., Wilkinson, R., & Lock, S. (2002). A single case study of conversation-focused therapy for a couple where one partner has aphasia. British Aphasiology Society Therapy Symposium Proceedings. London: British Aphasiology Society, 1-9.

Carroll, C., Patterson, M., Wood, S. Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. Implementation Science, 40 (2). DOI:10.1186/1748-5908-2-40

Chan, E.K.H., O'Neill, I., McKenzie, M., Love, A., & Kissane, D.W. (2004). What works for therapists conducting family meetings: treatment integrity in family-focused grief therapy during palliative care and bereavement. Journal of Pain and Symptom Management, 27 (6), 502-512.

Cherney, L.R., Simmons-Mackie, N., Raymer, A., Armstrong, E., & Holland, A. (2013). Systematic review of communication partner training in aphasia: methodological quality. International Journal of Speech-Language Pathology, 15 (5), 535-545.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. British Medical Journal (337:a1655), 979-983. DOI: http://dx.doi.org/10.1136/bmj.a1655

Di Rezze, B., Law, M., Gorter, J.W., Eva, K., & Pollock, N. (2012). A Narrative Review of Generic Intervention Fidelity Measures. Physical & Occupational Therapy in Pediatrics, 32 (4), 430-446.

Godfrey, E., Chalder, T., Ridsdale, L., Seed, P., & Ogden, J. (2007). Investigating the ‹active ingredients› of cognitive behaviour therapy and counselling for patients with chronic fatigue in primary care: Developing a new process measure to assess treatment fidelity and predict outcome. British Journal of Clinical Psychology, 46, 253-272.

Goodglass, H., & Kaplan, E. (2001). Boston Diagnostic Aphasia Examination. Austin Texas: Pro-Ed.

Originalbeitrag: Investigating treatment fidelity in a conversation-based aphasia therapy

Hallgren, K.A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutorials in Quantitative Methods for Psychology, 8 (1), 23-34.

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. Implementation Science, 67 (5). DOI:10.1186/1748-5908-5-67

Hasson, H., Blomberg, S., & Dunér, A. (2012). Fidelity and moderating factors in complex interventions: a case study of a continuum of care program for frail elderly people in health and social care. Implementation Science, 23 (7).

Hennessey, M.L., & Rumrill, P.D. (2003). Treatment fidelity in rehabilitation research. Journal of Vocational Rehabilitation, 19, 123-126.

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, 279-284.

Hogue, A., Liddle, H.A., & Rowe, C. (1996). Treatment adherence process research in family therapy: a rationale and some practical guidelines. Psychotherapy, 33 (2), 332-345.

Holland, A.L. (1991). Pragmatic Aspects of Intervention in Aphasia. Journal of Neurolinguistics, 6 (2), 197-211.

Holland, A.L. (1997). Pragmatic assessment and treatment for aphasia. In G.L. Wallace (Ed.), Adult Aphasia Rehabilitation (pp.161-174). Boston: Butterworth-Heinemann.

Huber, W., Poeck, K., & Springer, L. (2006). Klinik und Rehabilitation der Aphasie. Stuttgart: Thieme.

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, 369-379.

Kagan, A. (1998a). Supported conversation for adults with aphasia: Methods and resources for training conversation partners. Aphasiology, 12 (9), 816–830.

Kagan, A. (1998b). Philosophical, practical and evaluative issues associated with «Supported Conversation for Adults with Aphasia»: a reply. Aphasiology, 12 (9), 851-864.

Kagan, A., & Gailey, G. (1993). Functional is not enough: training conversation partners of aphasic adults. In A. Holland & M.M. Forbes (Eds.), Aphasia treatment: world perspectives (pp.199-226). London: Chapman & Hall.

Kiran, S., & Thompson, C.K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar atypicality. Journal of Speech, Language and Hearing Research, 46, 608-622.

Kolb, D.A. (1984). Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice Hall.

Lewinsohn, P.M., Clarke, G.N., Hops, H., & Andrews, J. (1990). Cognitive-Behavioral Treatment for Depressed Adolescents. Behavior Therapy, 21, 385-401.

Lichstein, K.L., Riedel, B.W., & Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. Advances in Behaviour Research and Therapy, 16 (1), 1-29.

Lock, S., Wilkinson, R., & Bryan, K. (2001a). Supporting Partners of People with Aphasia in Relationships and Conversation (SPPARC): A Resource Pack. Bicester, UK: Speechmark.

Lock, S., Wilkinson, R., Bryan, K., Maxim, J., Edmundson, A., Bruce, C., & Moir, D. (2001b). Supporting Partners of People with Aphasia in Relationships and Conversation (SPPARC). International Journal of Language & Communication Disorders, 36 (1), 25-30.

Lyon, J.G., Cariski, D., Keisler, L., Rosenbek, J., Levine, R., Kumpula, J., Ryff, C., Coyne,S., & Blanc, M. (1997). Communication partners: Enhancing participation in life and communication for adults with aphasia in natural settings. Aphasiology, 11 (7), 693-708.

McVicker, S., Parr, S., Pound, C., & Duchan, J. (2009). The Communication Partner Scheme: A project to develop long-term, low-cost access to conversation for people living with aphasia. Aphasiology, 23 (1), 52-71.

Moncher, F.J., & Prinz, R.J. (1991). Treatment fidelity in outcome studies. Clinical Psychological Review, 11, 247-266.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: development, measurement and validation. American Journal of Evaluation, 24, 315–340.

Nelson, M.C., Cordray, D.S., Hulleman, C.S., Darrow, C.L., & Sommer, E.C. (2012). A Procedure for Assessing Intervention Fidelity in Experiments Testing Educational and Behavioral Interventions. The Journal of Behavioral Health Services & Research, 1-22.

Odom, S.L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., Horn, E., Palmer, S., & Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. Early Childhood Research Quarterly, 25, 314-328.

Papathanasiou, I., Coppens, P., & Potagas, C. (2013). Aphasia and Related Neurogenetic Communication Disorders. Burlington: Jones & Bartlett Learning.

Resnick, B., Bellg, A.J., Borrelli, B., De Francesco, C., Breger, R., Hecht, J., et al. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. Annals of Behavioral Medicine, 29 (2), 46-54.

Saldert, C., Backman, E., & Hartelius, L. (2013). Conversation partner training with spouses of persons with aphasia: A pilot study using a protocol to trace relevant characteristics. Aphasiology, 27 (3), 271-292.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86 (2), 420-428.

Simmons-Mackie, N., Raymer, A., Armstrong, E., Holland, A., & Cherney, L.R. (2010). Communication Partner Training in Aphasia: A Systematic Review. Archives of Physical Medicine and Rehabilitation, 91, 1814-1837.

Swinburn, K., McVicker, S., & Pearce-Willis, E. (2011). Living with aphasia for life: What can community-based voluntary organizations offer? Aphasie und verwandte Gebiete, 2, 36-46.

Waltz, J., Addis, M.E., Koerner, K., & Jacobson, N.S. (1993). Testing the Integrity of a Psychotherapy Protocol: Assessment of Adherence and Competence. Journal of Consulting and Clinical Psychology, 61 (4), 620-630.

Whyte, J., & Hart, T. (2003). It's more than a black box; it's a Russian doll: Defining rehabilitation treatments. American Journal of Physical Medicine & Rehabilitation, 82 (8), 639-652.

Wielaert, S., & Wilkinson, R. (2012). Partners van Afasiepatienten Conversatie Training (PACT). [Conversation Training for Partners of People with Aphasia] Houten: Bohn Stafleu van Loghum.

Wilkinson, R., & Wielaert, S. (2012). Rehabilitation Targeted at Everyday Communication: Can We Change the Talk of People With Aphasia and Their Significant Others Within Conversation? Archives of Physical Medicine and Rehabilitation, 93, Suppl 1, 70-76.

Wilkinson, R., Bryan, K., Lock, S., & Sage, K. (2010). Implementing and evaluating aphasia therapy targeted at couples' conversations: A single case study. Aphasiology, 24 (6), 869-886.

Yeaton, W.H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity and effectiveness. Journal of Consulting and Clinical Psychology, 49 (2), 156-167.