AJSLP

Research Article

Developing, Implementing, and Improving Assessment and Treatment Fidelity in Clinical Aphasia Research

Leigh Ann Spell,[a] Jessica D. Richardson,[b] Alexandra Basilakos,[a] Brielle C. Stark,[c,d] Abeba Teklehaimanot,[e] Argye E. Hillis,[f,g] and Julius Fridriksson[a]

Purpose: The purpose of this study was to describe the development and implementation of a fidelity program for an ongoing, multifacility, aphasia intervention study and to explain how initial fidelity measures are being used to improve study integrity.
Method: A Clinical Core team developed and incorporated a fidelity plan in this study. The aims of the Clinical Core team were to (a) supervise data collection and data management at each clinical site, (b) optimize and monitor assessment fidelity, and (c) optimize and monitor treatment fidelity. Preliminary data are being used to guide ongoing efforts to preserve and improve the fidelity of this intervention study.
Results: Preliminary results show that specific recruitment strategies help to improve appropriate referrals and that accommodations to participants and their families help to maintain excellent retention. A streamlined and centralized training program assures the reliability of assessors and raters for the study's assessment and treatment protocols. Ongoing monitoring of both assessment and treatment tasks helps to maintain study integrity. Less-than-optimal interrater reliability data for the raters of some of the discourse measures guided the Clinical Core team to address the training and coding inconsistencies in a timely manner.
Conclusions: The creation of a Clinical Core team is instrumental in developing and implementing a fidelity plan for improved assessment and treatment fidelity. Intentional planning and assignment of study staff to implement and monitor ongoing fidelity measures assures that clinical data are reliable and valid. Ongoing review of the plan shows areas of strengths and weaknesses for continuing adjustments and improvement of study fidelity.

The usefulness of speech-language treatment to improve aphasia is supported by substantial evidence—people with aphasia who participate in speech-language treatment demonstrate improved outcomes compared to those who do not receive treatment (Brady et al., 2016; Holland et al., 1996; Robey, 1994, 1998). Nonetheless, individual response to treatment is highly variable, and little is known about how factors related to assessment, treatment, and participant characteristics combine to induce treatment responsiveness. With regard to assessment, different assessment foci can contribute to the mixed results (Bothe & Richardson, 2011; Cherney et al., 2011; Elsner et al., 2015; Worrall et al., 2011). For example, if discrete language abilities are treated and assessed, improvement in those language abilities (relative to a control condition) may be considered to support treatment efficacy. An additional view is that treatment is only efficacious if there is a demonstration of treatment-induced improvement in communication activities or life participation. Notably, different assessment administration procedures, particularly those involving the participant response and resultant score (item time limits, multidimensional scoring, self-corrected responses counted as accurate, etc.), can also lead to a participant being labeled as a *responder* with one set of procedures and as a *nonresponder* with another.

[a]Center for the Study of Aphasia Recovery, University of South Carolina, Columbia
[b]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque
[c]Department of Speech and Hearing Sciences, Indiana University Bloomington
[d]Program in Neuroscience, Indiana University Bloomington
[e]Department of Public Health Sciences, Medical University of South Carolina, Charleston
[f]Department of Physical Medicine and Rehabilitation, School of Medicine, Johns Hopkins University, Baltimore, MD
[g]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD
Correspondence to Leigh Ann Spell: spelll@mailbox.sc.edu
Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Anastasia Raymer
Received May 17, 2019
Revision received July 9, 2019
Accepted October 21, 2019
https://doi.org/10.1044/2019_AJSLP-19-00126

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Furthermore, regarding treatment, many factors (treatment elements, dose, etc.) are often studied, but findings are not straightforward. Although the "more treatment is better" view is generally supported (Basso, 2005; Bhogal et al., 2003; Brady et al., 2016; Cherney et al., 2008; Robey, 1998), it is unclear which treatment elements and targets should be emphasized (Barthel et al., 2008; Brady et al., 2016; Robey, 1998) and which outcome domains (e.g., body functions, activity limitations, participation limitations) benefit from intensive treatment schedules (Cherney et al., 2011). Finally, numerous participant characteristics, including stroke-related (e.g., lesion site and size, type of stroke, etc.), demographic, and neurophysiological variables, have been the focus of many investigations, often with conflicting results (Fridriksson et al., 2012; Kertesz & McCabe, 1977; Lazar et al., 2008; Plowman et al., 2012; Robey, 1998; Szaflarski et al., 2013).

Such variability in assessments, treatments, and participant characteristics leads to inconsistent results in the aphasia treatment literature, and researchers and practitioners have a difficult time identifying the best evidence to apply to their sample. The aforementioned examples are often predetermined during study design—the assessment measures, the treatment approaches, and the participant sample are selected to address the specific aims of each study. There are ample guidelines for reporting on study design and assessing the quality of that design. What receives comparatively little attention are the procedures, and reporting thereof, to ensure the study was implemented as designed, from initial recruitment procedures all the way to final data analysis. Once a study has been designed with the appropriate methodological rigor suitable for study aims, it is critical that the study is implemented in a way that mitigates threats to that study's validity. In the following sections, we will discuss types of validity especially related to adherence to study protocols as well as sampling, treatment and assessment fidelity in aphasia, and study recruitment and retention.

Study Validity

The level of confidence one can have in study results relates directly to study validity, or how closely a study's inference approximates the truth, and measures what it states that it measures (Shadish et al., 2002). One type of validity is statistical conclusion validity and relates to inferences about the presence of a relationship between two variables and the strength of that relationship. It can be threatened by low statistical power, measurement error, unreliable treatment implementation, violation of statistical test assumptions, and other sources of variance introduced into the experimental setting (Shadish et al., 2002). These threats can increase the chance of Type I or Type II error, or of an additional error (sometimes referred to as *Type III*) of concluding significance or nonsignificance, when in fact, the tests or the treatment protocols were not correctly administered (Bellg et al., 2004; Hinckley & Douglas, 2013; Nigg et al., 2002; Sánchez et al., 2007). For example, across the behavioral and health science literature, treatments or curricula administered with high fidelity generally result in larger effect sizes, but positive and detectable outcomes can still occur following those administered with low fidelity, lending unearned support for treatment elements (e.g., Girolametto et al., 2012; Hansen et al., 1991; Milburn et al., 2015; Solomon et al., 2000). Additionally, inflated effect sizes are observed when subjective measures are not rated without bias and/or when raters are not properly trained on scoring procedures, again lending strong but unearned support to the intervention (e.g., Hróbjartsson et al., 2013; Stitt et al., 2003). Another type of validity known as internal validity relates to whether or not causation can be inferred from the statistical conclusions. Internal validity is vulnerable to similar threats as statistical conclusion validity as well as participant selection and attrition, and various assessment factors (Richardson et al., 2016; Shadish et al., 2002). In other words, poor study implementation could be an additional contributor to the historically heterogeneous results (i.e., noise) in aphasia treatment research. In the absence of implementation planning and monitoring, investigators cannot confidently determine whether or not results (significant or nonsignificant) were caused by the targeted independent variable or were due to other random factors introduced because of poor sampling and/or poor protocol fidelity because the clinician "drifted" from the protocol (inconsistent/incorrect administration of cueing hierarchy, gradual changes over time to scaffolding and/or feedback, etc.) or "contaminated" the protocol (adding or omitting elements, including elements from other treatment protocols, etc.; Bellg et al., 2004).

Treatment Fidelity

Most implementation discussions have so far focused on treatment fidelity or how well the essential treatment elements were delivered as intended and were distinguishable from comparison conditions (Bellg et al., 2004; Hinckley & Douglas, 2013; Gearing et al., 2011). The National Institutes of Health established a treatment fidelity workgroup within the Behavior Change Consortium in recognition of poor monitoring of treatment fidelity in behavioral research and the impact on study validity. This workgroup was tasked with defining treatment fidelity and offering guidelines for researchers. According to the Behavior Change Consortium workgroup, establishing treatment fidelity should address the following five components (Bellg et al., 2004): (a) study design—the study should be designed appropriately so that hypotheses can be tested and inferences are valid, (b) training—training procedures should be standardized across clinicians, (c) treatment delivery—treatment should be monitored to ensure it is delivered as intended, (d) treatment receipt—the participant should understand the treatment procedures and demonstrate utilization of them within experimental sessions, and (e) treatment enactment—the participant should utilize behaviors targeted in treatment in real-world settings.

Since the dissemination of these guidelines, there have been investigations of the relationship between procedural fidelity and treatment outcomes. A common finding is that studies taking steps to ensure high fidelity increases the power to detect effects that may have otherwise been obscured by variance, with different dimensions of fidelity serving as significant moderators of effect size, depending on the nature of the intervention (Claridge, 2014; Hansen et al., 1991; Koehler et al., 2013; Maxfield & Hyer, 2002). For example, in general psychology research, several meta-analyses demonstrate that studies taking steps to ensure treatment fidelity have larger effect sizes (two to three times higher) for treatment outcomes compared to those studies that do not (Durlak & DuPre, 2008). Additionally, school psychology researchers have revealed that, as the degree of treatment fidelity increases, the rates of positive outcomes also increase (Gresham et al., 1993). Thus, implementation can be a determinant of success.

Hinckley and Douglas (2013) examined the reporting of treatment fidelity in aphasia treatment studies (reviewing 149 aphasia treatment articles published between 2002 and 2011), revealing that half of the studies reported replicable treatment methods, but only 21 out of 149 (14%) of the studies specifically described methods for establishing or monitoring treatment fidelity. Twenty of those studies utilized only a single treatment fidelity method—either (a) monitoring adherence to treatment protocol, (b) supervising treatment sessions, (c) training manual utilization, or (d) role playing. A single study reviewed utilized more than one treatment fidelity method, combining training manual utilization with monitoring of treatment protocol adherence.

Since Hinckley and Douglas' (2013) review, more attention has been paid to fidelity in aphasia intervention studies. Several studies have incorporated video-recorded treatment sessions in order for a fidelity rater to evaluate a clinician's implementation of a treatment protocol (Godecke et al., 2015; Kladouchou et al., 2017; Salis et al., 2017; Volkmer et al., 2018; Worrall et al., 2016). Other studies have reported on the use of fidelity checklists to monitor adherence to home intervention programs by both family members (Behn et al., 2018) and persons with aphasia (PWAs; Ball et al., 2018). In a recent review, Brogan et al. (2019) examined randomized control trials of aphasia treatment published between 2012 and 2017, revealing that, although 37 out of 42 (88%) of the randomized control trials addressed the study design aspect of treatment fidelity, only nine out of 42 (21%) explicitly discussed treatment fidelity processes. Although more authors seem to be addressing treatment in the initial design of their studies, aphasia research is still lacking in a consistent, explicit description of other ongoing treatment fidelity processes. Overall, adequate and multidimensional treatment fidelity monitoring and reporting is still more of an exception rather than the norm.

Assessment Fidelity

Compared to treatment fidelity, similar guidelines to ensure adherence to an assessment protocol have not been established by the consortium. The general impression is that selection of tests with good reliability and validity, and perhaps performing rater reliability checks within the study, is enough to ensure assessment fidelity. However, just as there can be clinician drift from treatment procedures and contamination of treatment procedures, assessor and/or rater drift and contamination are just as likely to occur, especially when administering lengthy assessment batteries that include tests with different administration procedures, complex scoring systems, and/or repeated administration over a long period of time. Richardson et al. (2016) reviewed 88 aphasia treatment studies published between 2010 and 2015 and examined the frequency with which researchers provided information about assessment fidelity components. Results showed that, of the 88 studies reviewed, only 57% provided any information regarding assessment fidelity. Of those studies that did describe assessment fidelity information, 37.5% reported on assessor reliability, 35.2% reported on assessor qualifications, 27.3% reported on assessor blinding, 6.8% reported on assessor training, and 4.5% reported on information regarding assessment instruments (Richardson et al., 2016). Recommendations to improve and monitor assessment fidelity were also provided and included predetermined assessor and rater training and qualifications, use of training manuals, video observation of administration and scoring methods, role play and monitoring of practice assessments and scoring with immediate feedback, booster training sessions for scoring and administration, adherence monitoring, and more (Richardson et al., 2016). Compared to treatment fidelity, assessment fidelity has received little attention, and the influence of assessment fidelity on power and effect sizes is not currently calculable but certainly suspected.

Study Sampling

Perhaps the most important aspect of data collection is ensuring that there are participants from whom to collect data and that the participants sampled represent the population in a way that promotes generalization. Despite this, guidelines for participant recruitment, enrollment, and retention plans and reporting are rare to nonexistent; though such plans would guard against deficient power, selection bias, and attrition, all threats to statistical conclusion and internal validity have been discussed previously (Shadish et al., 2002). Even reporting of bare minimum design elements related to allocation and its concealment is not consistent in the aphasia treatment literature, so that the majority of studies included in a recent Cochrane review either had a high risk of bias or an unclear risk of bias, unclear because information was not reported to allow authors to determine bias (Brady et al., 2016).

The purpose of this study was to describe the development and monitoring of an implementation plan for an ongoing, multifacility, aphasia intervention study and to provide additional recommendations stemming from our efforts to maintain and improve study integrity.

Method

To assure quality data collection and management as well as assessment and treatment fidelity in an ongoing National Institute on Deafness and Other Communication Disorders project (P50 DC014664) coordinated by the University of South Carolina (UofSC) Center for the Treatment of Aphasia Recovery (C-STAR), a Clinical Core team was created to make sure that all assessors, raters, and clinicians adhere to the study protocols. The Clinical Core team includes neurologists, speech-language pathologists (SLPs), neuroscientists, postdoctoral fellows, and graduate students located at the UofSC, the Medical University of South Carolina (MUSC), and Johns Hopkins University. This team met weekly for the first year of the grant to discuss study setup and barriers but over time, in Year 2, reduced meeting frequency to monthly meetings to update progress and consult on any questions the team members may have. All fidelity strategies were implemented from the beginning of the study, and a few strategies have been adjusted as needed. For example, when traditional recruitment methods were not yielding the number of participants expected, we shifted from advertising to a professional referral program. Another example of an adjustment that we made in regard to assessment fidelity was when our SLPs felt like they needed more training on a specific assessment (Apraxia of Speech Rating Scale [ASRS]; Strand et al., 2014) early in the study; we scheduled additional training with the expert on that assessment and scheduled weekly practice group scoring meetings to ensure consistency. The aims of the Clinical Core team are (a) to supervise data collection and data management at each clinical site, (b) to optimize and monitor assessment fidelity, and (c) to optimize and monitor treatment fidelity. Table 1 shows the specific areas addressed with each aim.

This article focused on one C-STAR project under the supervision of the Clinical Core team, the Predicting Outcomes of Language Rehabilitation (POLAR). This project involves the ongoing recruitment of 120 stroke survivors with aphasia and 30 stroke survivors without aphasia; administration of an extensive list of assessments for baseline testing by four, American Speech-Language-Hearing Association–certified and state-licensed, research SLPs at two facilities (UofSC and MUSC); a 6-week, daily intervention program including both semantic and phonological-focused treatment tasks for PWAs; and follow-up testing with outcome measures. The total length of time that treated participants are enrolled in the study is 42 weeks, with 6 weeks total of daily treatment and follow-up testing taking place at 1 month posttreatment and 6 months posttreatment. Eleven graduate students in the Master's in Speech Pathology program at the UofSC serve as raters for all outcome measures. At this time, we have enrolled 81 participants in the study.

Aim 1: To Supervise Data Collection and Data Management at Each Clinical Site

The first aim of the Clinical Core team has three main components: to make sure that all team members are adequately trained at baseline, that they are able to recruit/retain appropriate participants, and that they are able to accurately collect and manage the data for these projects.

Baseline Training

An ongoing activity toward this aim is baseline training of new SLPs, who serve as assessors as well as treatment providers and raters for some of the assessments. Baseline training also takes place with graduate students who serve as raters for the naming and discourse assessments. Baseline training includes human subjects training (Collaborative Institutional Training Initiative (CITI) Program, n.d.), specific protocol review, informed consent training with aphasia friendly materials, assessment review (administration/scoring), and treatment review. The Clinical Core team has streamlined all training of SLPs and raters with detailed protocol manuals and video-recorded training sessions for both assessment and treatment administration as well as for outcome rater (graduate student) training. Clinicians and raters go through a detailed, two-day training regimen with immediate follow-up observations and reviews of performance to make sure that they demonstrate competence with all tasks.

Measurement of the efficiency of training is done through verbal and written feedback from the SLPs and raters after they complete baseline training. The Clinical Core Coordinator observes SLPs during their initial assessment and treatment sessions with participants and provides verbal and/or written feedback after each session observed. Second-year, experienced graduate students serve as second

Table 1. Fidelity aims and specific areas addressed.

Aim	Description	Areas addressed
Aim 1	To supervise data collection and data management at each clinical site	Clinician and rater baseline training Recruitment and retention of participants Data capture and management
Aim 2	To optimize and monitor assessment fidelity	Ongoing clinician assessment training and support Monitoring of assessment delivery Ongoing rater training and support
Aim 3	To optimize and monitor treatment fidelity	Ongoing clinician treatment training and support Monitoring of treatment delivery Monitoring of treatment receipt

raters for first-year students immediately after their training and provide verbal and/or written feedback on their coding assignments.

Recruitment and Retention

The second ongoing activity toward Aim 1 is recruitment and retention of participants. Recruitment is monitored through the use of the Recruitment Index, which is the number of days required to recruit an analyzable participant at one site (Rojavin, 2005). The number of participants enrolled each month is also monitored. Initial recruitment relied on contacting participants from previous studies in our labs, newspaper/radio/television advertising, and professional referrals. Current recruitment strategies focus more on professional referrals and community outreach programs with advertising through our websites: the Aphasia Lab at the UofSC (n.d.; https://web.asph.sc.edu/aphasia/) and the C-STAR (https://cstar.sc.edu/), as well as through social media, specifically our Aphasia Lab-USC (n.d.) Facebook page (n.d.; https://www.facebook.com/StrokeRecoveryProject/).

Recruitment and retention activities have included educational luncheons, tailgating before UofSC football games, a community leadership class, a monthly lunch group, a pen pal activity to connect with PWAs in North Carolina, a blog for participants describing publications in an aphasia-friendly way, podcasts, and a drama group. Retention at all sites has been addressed by recruiting motivated participants and accommodating their needs. A Clinical Coordinator at each study site completes phone intake interviews with potential participants to explain eligibility criteria, the course of the study (including the significant time commitment), and to answer questions from participants and their family members. Lodging is provided for out-of-town participants. Transportation assistance is also available. Clinicians go to participants' homes when participants are not able to come to the study sites. Scheduling is set up around participants' and their families' needs. We send quarterly newsletters to all current and past participants to keep them up to date on events as well as summaries of recent published research. At the UofSC, participants are encouraged to participate in weekly stroke recovery groups, which are offered at no cost, even after they have completed a study. A family support group is also in development at the UofSC. At the MUSC, a monthly stroke support group is offered to participants and they are encouraged to attend.

Data Capture and Management

We have worked with faculty and staff from the Data Coordination Unit (DCU) at the MUSC to set up a data entry and storage system for all baseline and outcome data. This web-based clinical trial data management system, Web Data Coordination Unit (WebDCU, n.d.), was developed to manage clinical trials (https://dcu.musc.edu/). Assessment and treatment apps have been developed so that video recordings of assessment and treatment tasks are uploaded automatically to Health Insurance Portability and Accountability Act–compliant Dropbox (at the UofSC) and box (at the MUSC) accounts as soon as they are administered. After assessments are scored by assessors and raters, these data are monitored and analyzed by Data Coordination Unit staff.

The stability of participants' performance is monitored by administering one of the naming assessments (Philadelphia Naming Test [PNT]; Roach et al., 1996) twice on two different days. In this assessment, participants are shown pictures of items and are asked to name them. Often times, participants will take multiple attempts to name the items. Coding protocol requires that we transcribe and code both their first attempt for each item and their last attempt for each item. To evaluate participant consistency in test performance, test–retest reliability for both first and last naming attempts on this assessment is calculated.

Aim 2: To Optimize and Monitor Assessment Fidelity

To ensure assessment fidelity, ongoing activities in this project include clinician assessment training and support, monitoring of assessment delivery, and rater training reliability and support.

SLP Assessment Training and Support

In-depth training with experienced assessors (SLPs) on each assessment tool with ongoing support is imperative for assessment fidelity. Our first ongoing activity toward this aim involves clinician qualifications and training. The six SLPs who have worked or are working on this project are experienced clinicians who are all certified by the American Speech-Language-Hearing Association and have South Carolina State Licensure. They range in years of professional experience from 4.5 to 37, and they have worked in a variety of settings including acute care hospitals, inpatient/outpatient rehabilitation hospitals, skilled nursing facilities, home health, assisted living facilities, private practices, and research institutions. As mentioned above, all initial assessment training has been centralized and streamlined so that all clinicians receive the same instruction on administration and scoring. For this project, assessors receive extensive training on administration and scoring of the following standardized assessments: the Western Aphasia Battery–Revised (Kertesz, 2007), the ASRS (Strand et al., 2014), the Pyramids and Palm Trees Test (Howard & Patterson, 1992), the Kissing and Dancing Test (Bak & Hodges, 2003), the Northwestern Assessment of Verbs and Sentences (Thompson, 2012), subtests of the Psycholinguistic Assessments of Language Processing in Aphasia (Kay et al., 2009), the Matrix Reasoning subtest of the Wechsler Adult Intelligence Scale (Wechsler, 2008), the Aphasia Communication Outcome Measure (Doyle et al., 2012), and subtests of the Temple Assessment of Language and Short-Term Memory in Aphasia (Martin et al., 2010). Although raters (graduate students) score the following assessments, the SLPs administer the PNT (Roach et al., 1996; Walker & Schwartz, 2012), the

Philadelphia Repetition Test (Dell et al., 2007), three discourse tasks (retelling of the Cinderella story, description of a four-panel picture sequence ["Broken Window"] story, and procedural description of how to make a peanut butter and jelly sandwich), and a naming test of 40 nouns and verbs treated in therapy activities. Ongoing consultation with assessment experts Jessica Richardson (fidelity, discourse), Dirk den Ouden (Northwestern Assessment of Verbs and Sentences), Brielle Stark (discourse), Grant Walker (naming), and Alexandra Basilakos (ASRS) ensures accurate administration, scoring, and analysis of assessment measures.

To monitor SLP training, the Clinical Core Coordinator observes each clinician giving each assessment initially and provides feedback on administration and scoring. When there are questions about an assessment measure, experts are consulted for more input.

Monitoring Assessment Delivery

The second ongoing activity toward this aim involves ongoing monitoring of assessor (clinician) delivery by the Clinical Core Coordinator, Leigh Ann Spell. She observes, documents, and provides feedback on assessment sessions for each incoming participant at the UofSC. She does the same for assessment sessions at the MUSC through video-recorded assessment sessions. The Clinical Core Coordinator keeps an assessment fidelity log for each SLP in which she records the date, the participant number, the assessment observed, and then records either "yes" or "no" as to whether or not the SLP (a) adheres to the assessment administration and (b) engages the participant in the assessment. If an SLP does not adhere to the assessment administration in any way or does not effectively engage the participant in the assessment, this information is shared with her immediately. This log allows the Clinical Core Coordinator to see if there are specific assessments that are more difficult to administer/score across all clinicians or if a specific clinician is having more difficulty than others with the assessment tasks. If there is a question about an assessment session, our assessment experts are consulted for more information and guided practice if necessary. To determine how well all SLPs are doing with assessment delivery, a percent average of all clinicians' performance is calculated through the number of items with a "yes" divided by the total number of sessions observed for each item.

Rater Training, Reliability, and Support

The third ongoing activity toward Aim 2 is rater training, reliability, and support. The Clinical Core Coordinator, along with the graduate students at the UofSC, is responsible for the transcription, coding, and scoring of all of the outcome measures. Raters are assigned to one of two teams: One focuses on naming outcomes, whereas the other focuses on discourse outcome measures. The naming team is specifically trained in the transcription and coding of video-recorded naming assessments based on the PNT scoring protocol (Moss Rehabilitation Research Institute, 2016). The discourse team is specifically trained in the transcription and coding of video-recorded discourse samples using Codes for the Human Analysis of Transcripts and Computerized Language Analysis tools (MacWhinney, 2000). More experienced raters (second-year graduate students) are paired with new raters (first-year graduate students) to provide transcription and coding feedback as well. When raters disagree on a code, they consult with the Clinical Core Coordinator or the discourse/naming consultants to resolve the disagreement.

To monitor rater reliability, inter- and intrarater reliability is analyzed on 10% of naming and discourse assessments scored at the end of each fall semester to check consistency (especially of new raters) and provide feedback to student raters. The Clinical Core Coordinator meets with students each semester with feedback on scoring updates and reliability data.

Aim 3: To Optimize and Monitor Treatment Fidelity

This aim is similar to Aim 2 but addresses treatment fidelity instead of assessment fidelity. It is important that all clinicians are providing treatment in the same way to assure valid and reliable outcome measures.

SLP Treatment Training and Support

An ongoing activity toward Aim 3 is clinician training for each treatment approach utilized in the study. Initial SLP training and monitoring have been described above in Aim 1—baseline training. To continue treatment training, SLPs meet monthly (or more often) to discuss questions about protocols. They also participate in online webinars and other professional development activities related to aphasia and apraxia of speech.

Monitoring Treatment Delivery and Receipt

To assure consistent treatment delivery and receipt, the Clinical Core Coordinator serves as the treatment fidelity rater. Treatment fidelity has been carefully monitored through face-to-face and video-recorded treatment sessions. This has been extremely helpful to make sure that all participants are receiving treatment in the same way. The Clinical Core Coordinator observes one treatment session for each therapy type (e.g., one phonologic session and one semantic session) and evaluates both treatment delivery and treatment receipt for each participant. As with assessment session observations, feedback is provided immediately to SLPs on their performance.

Monitoring of treatment fidelity (delivery and receipt) involves observation of approximately 10% of treatment sessions by the Clinical Core Coordinator. She keeps a treatment fidelity log for each SLP, which includes the date, participant number, type of treatment observed (semantic of phonologic), and then records either "yes" or "no" as to whether or not the SLP (a) adhered to treatment procedures and (b) enacted treatment and engaged the participant in tasks. In this log, treatment receipt for the participant is also measured by recording "yes" or "no" as to whether or not the participant (a) was engaged, (b) understood the tasks, and (c) consistently attempted to perform the tasks.

To determine how well all SLPs are doing with treatment delivery and receipt, a percent average of all clinicians' performance is calculated through the number of items with a "yes" divided by the total number of sessions observed for each item.

Preliminary Results

Aim 1: To Supervise Data Collection and Data Management at Each Clinical Site

Baseline Training

Baseline training has taken place whenever a new clinician or graduate student joins the project. Verbal and written feedback from new project members has included things like a request for a formal mentor or "scoring buddy" (from student raters), more professional development training (from SLPs and student raters), and more opportunities to observe clinicians administering treatment (from SLPs). All of these suggestions from student raters and SLPs are considered and implemented as needed to improve the baseline training experience.

We have had positive verbal feedback from student raters on using the mentoring model in their first semester of training. Having second-year students be second raters to first year students has increased students' confidence and ability to get "up to speed" on their scoring tasks as they are being trained. We have also provided professional development training for student raters. Every student rater has completed the *Introduction to Supported Conversation for Adults with Aphasia* webinar with the Aphasia Institute and is always invited to attend our bimonthly C-STAR lecture series.

Our SLPs have provided positive verbal feedback on increasing observation opportunities of experienced clinicians during both assessment and treatment sessions. Clinicians can observe either live or video-recorded sessions. We have also provided supplemental professional development opportunities through webinars, conferences, and our C-STAR lectures.

Recruitment and Retention

We have seen an increase in participant recruitment from Year 1 to Year 3 of the C-STAR project. The average number of enrolled participants per month has increased from an average of 1.67 participants per month in Year 1 of the grant to an average of 2.33 participants per month in Year 3 of the grant. The Recruitment Index has decreased from 32.0 days in Year 1 to 24.38 days in Year 3. Shifting from general advertising strategies to a specific, targeted referral program seems to be more effective in identifying and recruiting our target population.

Retention has been excellent throughout the study with a retention rate of 95%. Of the four people who have withdrawn from the study, two have withdrawn due to medical complications, one due to us not being able to contact him for his 6-month follow-up visit after he moved, and one due to the participant's frustration with the treatment tasks. Continuous communication with participants and provision of support groups and educational/social activities seem to help us retain participants. Although we are not sure which activities work best or whether all of these strategies are needed to retain participants, we have chosen to err on the side of using all of the techniques to make sure that we maintain retention.

Data Capture and Management

Staff from the MUSC's WebDCU provides feedback monthly on outstanding data or data entries that need clarification. SLPs and student raters reply to these requests monthly. As mentioned above, the stability of participants' performance is monitored by administering one of the naming assessments (PNT) twice on two different days. Test–retest reliability of the PNT for participants' first attempts at naming items is good–excellent, with intraclass correlations (ICCs) ranging from .62 to .95. The ICC for correctly named items at first attempt had excellent agreement (.95), but participants were generally more variable in some types of paraphasic errors like mixed errors (.62) and semantic errors (.64). Participants were more variable on their last attempts at naming items, with ICCs ranging from fair to excellent (ICCs from .48 to .80).

Aim 2: To Optimize and Monitor Assessment Fidelity

SLP Assessment Training and Support

If a clinician has a question about a specific assessment, this is monitored and discussed in our monthly Clinical Core team meetings. For example, early in the study, some SLPs were unsure about the ratings used with the ASRS (Strand et al., 2014). In response to this feedback, the expert consultant (Alexandra Basilakos) arranged for some reliability scoring sessions with all clinicians to make sure that everyone was consistent in how they were rating the different subtests of the ASRS. Another example involved administration of the discourse assessments. One SLP was not sure about what to do when a participant did not produce any language for a discourse task. This was discussed at the next SLP meeting, and strategies were discussed for encouraging participants with severe expressive language deficits to verbalize as much as they could on the discourse tasks.

Monitoring Assessment Delivery

Clinician assessment delivery is measured through direct observation or observation of video-recorded assessment sessions. The assessment fidelity rater keeps an assessment fidelity log for each SLP, which includes whether or not the SLP adheres to the assessment administration protocol and whether or not she enacts the assessment and engages the participant. Approximately 10% of all sessions are observed. Review of these logs show that, in observed sessions from Year 1 through Year 3 of the study, SLPs have adhered to the assessment protocols with 94% accuracy and have enacted the assessment and engaged the participant

with 100% accuracy. Since this adherence to assessment protocol has been consistently high for all clinicians throughout the study, it could be an area where fewer observations are necessary.

Rater Training, Reliability, and Support

For rater reliability, intra- and interrater reliability using ICCs with two-way mixed model (ICC[2,1]) for absolute agreement was used. This type of model is appropriate for when each outcome is rated by each rater, and raters are considered representative of a larger sample of similar raters. For the primary outcome measure of naming (PNT), interrater agreement (.76–.99) and intrarater agreement (.98–.99) have been excellent. For discourse, inter- and intrarater reliability have largely been excellent (0.82–0.98) over the course of the study, but some aspects of discourse have proved more difficult to obtain strong reliability within and across raters. In large part, the difficulty has arisen in the identification of phonemic and semantic paraphasias in the discourse tasks, where ICC values have ranged from 0.22 to 0.65 through the study. Another area that seemed to affect raters' reliability was utterance segmentation. When this was identified, our discourse consultants provided more specific guidelines for students to improve utterance segmentation. We believe these procedures help further guide us in our rater training and ongoing support.

Aim 3: To Optimize and Monitor Treatment Fidelity

SLP Treatment Training and Support

As with assessment training and support, questions or concerns about treatment protocols after initial training are addressed at our monthly SLP meetings. One example of this was when an SLP was not sure about the amount of cuing allowed with a specific semantic treatment approach. This was discussed at the next SLP meeting, and all SLPs agreed and clarified the protocol for that specific task. In response to continuing technology issues with the laptops that we use for some treatment tasks, a programming specialist was hired to help with our study apps and with other studies' needs.

Monitoring Treatment Delivery and Receipt

We have seen very good adherence to treatment protocols by clinicians. Just as with assessment fidelity, this has included informal feedback from SLPs through monthly meetings (as noted above) and observation of therapy sessions by the Clinical Core Coordinator. Treatment fidelity logs (as described in the Method section) showed the following: In observed sessions, SLPs have adhered to the treatment protocols with 83% accuracy and have enacted the treatment and engaged the participant in tasks with 100% accuracy. Examples of deviations from the protocol included presenting the steps of a task slightly out of order or spending more than the allotted time on a specific task. SLPs received immediate written or oral feedback on observed sessions. Treatment receipt is also observed during sessions. Review of these logs showed that, in observed sessions, participants were engaged in 100% of sessions, they understood tasks in 82% of sessions, and they consistently attempted to perform tasks in 95% of sessions. At times, the participant's aphasia type and severity affected treatment receipt (e.g., understanding the task), but the Clinical Core Coordinator still examines whether the SLP still engages the participant, tries to help him/her comprehend what is being asked, and motivates the participant to consistently perform the task. Since this adherence to treatment protocol has been consistently high for all clinicians throughout the study, it could be an area where fewer observations are warranted, especially after an SLP gains more experience with a variety of different participants.

Discussion

Treatment and assessment fidelity are critical components for effective translation of clinical research into evidence-based practice (Breitenstein et al., 2012; Gearing et al., 2011; Mowbray et al., 2003). Although more investigators are incorporating fidelity components into their aphasia studies, most do not describe a systematic way of addressing assessment or treatment fidelity (Brogan et al., 2019; Hinckley & Douglas, 2013; Richardson et al., 2016). In this study, the investigators were intentional in setting up an explicit fidelity plan to develop, implement, and improve both assessment and treatment fidelity throughout the course of the study. A Clinical Core team consisting of clinical research staff was organized to reach this goal. For each project aim, measurement of fidelity is essential in assuring the quality of the behavioral data collected.

Aim 1: To Supervise Data Collection and Data Management at Each Clinical Site

The investigators learned early in the project that centralized baseline training allowed all personnel to receive the same training and feedback for more accurate and consistent data collection. This is consistent with the findings of Shadish et al. (2002), who reported that better training of assessors improves study validity and increases power. An early occurrence that was very helpful was that SLPs in the POLAR project developed a detailed protocol manual for each assessment and treatment task. This required all clinicians to take ownership of the protocol which, in turn, contributed to "buy in" to the process and an increased sensitivity to adherence to procedures (Richardson et al., 2016). Another aspect of the initial training for clinicians was a thorough review of the different treatment tasks used in the study and the rationale for using each task. Roth and Pilling (2008) stated that initial training should foster this "meta-competence" of understanding the theories and rationales behind treatment components to increase providers' ability to be flexible with different levels of participants while adhering to the treatment protocol. Another helpful training technique was video-recording all initial

and follow-up trainings. This is convenient for new personnel (especially graduate students who have varying schedules) and also readily allows experienced assessors and raters to refer back to the training as needed.

Recruitment and retention of appropriate participants is essential to the success of any human research study (Gul & Ali, 2010; Patel et al., 2003; Rojavin, 2005). Gul and Ali (2010) found that tracking successes and challenges faced throughout a study helps to identify and adjust recruitment and retention strategies. Early on in this project, we implemented traditional advertising strategies like running newspaper, radio, and television ads. Although these ads generated many calls, analysis of the number of eligible participants that were actually generated from each ad showed that the return on investment for each ad was very low. Since we found our best referrals came from other rehabilitation professionals, in the third year of the grant, we have shifted focus from general advertising to a professional referral program, which has helped us recruit more appropriate participants at a much lower cost. Monitoring successful retention strategies has also been beneficial in helping participants complete this 42-week study. Many participants initially take part in a clinical treatment study hoping that it will benefit them directly or help others (Miller et al., 1998; Patel et al., 2003). Maintaining participation in a study is sometimes difficult, however. Helping participants overcome potential barriers such as communication, cost, travel, time, and limited social support makes it easier for them to complete the study (Gul & Ali, 2010; Patel et al., 2003). Retention strategies that have been effective for this study include implementing frequent communication with participants and caregivers, providing lodging (which allows pets) for our out-of-town participants, providing transportation assistance, providing other optional programs for participants (e.g., group therapy, social and educational activities) and accommodating participants' schedules. To learn even more from our participants on their research experience, we are in the process of implementing an anonymous satisfaction survey that they will complete with their caregivers after they finish the treatment phase of our study.

Data capture and management has been one of the many strengths of the POLAR. Although we had some minor difficulties with video-recording applications and storing the large number of videos that we gather initially, hiring staff to address technology issues has decreased these problems. Working with WebDCU has helped to maintain the integrity of the data collected and assure accurate data analysis. See Table 2 for the strategies that we found to be the most efficient with data capture and management.

Aim 2: To Optimize and Monitor Assessment Fidelity

As mentioned previously, ongoing clinician assessment training and support takes place through monthly Clinical Core and SLP meetings. Clinicians also utilize peer review of scoring when necessary. This is accomplished by having an SLP serve as a second rater for an assessment, and the two scoring forms are compared. Discrepancies are discussed between clinicians, and if forced agreement cannot be reached, the discrepancy in scoring is brought to the larger group for further discussion. Monitoring of assessment delivery is completed by the Clinical Core Coordinator either face-to-face or via video recording. Use of a checklist for targeted behaviors (e.g., adherence to the assessment protocols, engagement of the participant), as has been recommended by Borrelli (2011), has been beneficial to track assessment delivery. A challenge encountered during both assessment and treatment fidelity monitoring has been the time required to complete these observations for all clinicians. Pereletchikova et al. (2007) concurred that assessing integrity can be expensive and time consuming.

One disadvantage of having graduate students as raters is the fact that we have them for only 2 years before they graduate. To address this, the Clinical Core team has implemented a mentoring program so that new raters always have adequate support. The Clinical Core Coordinator completes all initial training, which includes viewing of the video-recording training and guided practice as well as feedback on all initial rating attempts. In addition to this initial training, new raters (first-year graduate students) are paired with experienced raters (second-year graduate students) who are second raters on all initial scoring assignments. All raters have access to the Clinical Core Coordinator (Spell) and to experts in the transcription and coding of naming (Walker) and discourse (Stark) to answer questions that cannot be answered by their peers.

Another challenge of having graduate students as raters is that they are less experienced than experienced clinicians in identifying communication errors. Some graduate students have training in linguistics or more advanced courses in speech, language, and hearing prior to arrival, but the majority does not. When poor interrater reliability was noted in the discourse team with the identification of paraphasias, additional training was provided to raters with a discourse expert (Stark). Identification of paraphasias is difficult because the targets of paraphasias can often be unclear due to the open-endedness of the prompt (e.g., Cinderella story has many parts). To address this inconsistency in paraphasia identification, the consultant (Stark) developed a flowchart to aid in decision making and met in-person or via videoconference with raters. In addition, a second-rater check system was implemented for better consistency in paraphasia identification in discourse. As recommended by Richardson et al. (2016), scoring and rating reliability should be consistently reported to establish efficacy. This has helped us determine in which areas raters need the most assistance and support for more accurate scoring. See Table 3 for the strategies that were most helpful in addressing assessment fidelity in this study.

Aim 3: To Optimize and Monitor Treatment Fidelity

According to Kelly et al. (2000), treatment fidelity criteria should include roles, qualifications, and activities

Table 2. Strategies for improving clinical research fidelity: data collection and management.

Area addressed	Strategies
Clinician and rater baseline training	Employ experienced clinicians with strong clinical skills
	Have experts in specific assessment/treatment protocols train clinicians and serve as consultants as needed
	Create detailed, written manuals for all assessment/treatment tasks
	Create detailed, written manuals for scoring/coding outcome measures
	Video-record all training to be used in multiple facilities with all clinicians and raters
Recruitment	Create and maintain an informative website
	Create an active social media presence
	Develop regularly occurring community outreach activities (e.g., monthly lunch group, support groups, drama club, tailgating)
	Develop a professional referral program with medical professionals (e.g., therapists, neurologists) who regularly see your target population
	Employ a diverse recruiting staff who can reach out to underrepresented, minority populations
Retention	Provide clear information about study requirements and schedule
	Accommodate participant schedules as much as possible
	Provide assistance with transportation and lodging
	Have one primary research staff member assigned to work with each participant
Data capture and management	Consult with a data management organization for collection, organization, and analysis
	Have technology support personnel for all electronic data gathering and data storage systems

of staff. In this study, it seems likely that having highly qualified SLPs with extensive clinical backgrounds has contributed to consistent implementation of treatment tasks. As mentioned previously, having clinicians develop the treatment protocols after a thorough review of the treatment approaches assured that all clinicians had a strong grasp of the therapy activities. Our strong retention strategies also contribute to having engaged, committed participants who consistently make every effort to complete treatment tasks. See Table 4 for strategies that were most helpful in addressing treatment fidelity in this study.

Limitations and Future Research

Since developing, implementing, and improving an assessment and treatment fidelity plan is a dynamic process, a few limitations have come to light during this project. First, although networking with rehabilitation professionals and completing community outreach activities have increased recruitment and retention of appropriate participants, a time-efficient and effective strategy of maintaining recruitment and reaching out to minority and lower socioeconomic groups is still in development. Sufficient time and staffing should be dedicated to recruitment (Gul & Ali, 2010) with the use of more interactive recruitment channels (e.g., telephone, interpersonal communication) versus passive recruitment channels (e.g., mass media; McDonald, 1999). Discussions to address these issues in the future have included collaborating with other labs on recruitment and hiring staff to focus almost exclusively on recruitment. Another future strategy is to have a more diverse recruiting staff and to reach out to areas of religious institutions where we can reach potential participants who we are missing in other venues.

Second, in this study, only one person (the Clinical Core Coordinator) has been tasked with evaluating clinician assessment and treatment fidelity. Other studies have shown that it is beneficial to have several sources of fidelity assessment such as using multiple raters, using peer raters, implementing self-assessment, and using participant input

Table 3. Strategies for improving clinical research fidelity: assessment fidelity.

Area addressed	Strategies
Ongoing clinician assessment training and support	Use assessments with strong validity and reliability
	Create detailed, written manuals for all assessment tasks
	Have experts in specific assessment protocols serve as consultants as needed
	Provide ongoing professional development support
	Hold regular clinician meetings to discuss assessment questions and concerns
Monitoring of assessment delivery	Have fidelity coordinators observe clinicians' assessment sessions regularly (either live or video-recorded) and provide feedback
	Have clinicians complete self-assessments and peer assessments for additional feedback
Ongoing rater training, reliability, and support	Create rater "teams" for rater specific training and peer mentoring
	Have experts in coding/scoring serve as consultants as needed
	Consistently check inter- and intrareliability of raters
	Hold regular rater meetings to provide feedback on reliability and to discuss questions and concerns

Table 4. Strategies for improving clinical research fidelity: treatment fidelity.

Area addressed	Strategies
Ongoing clinician treatment training and support	Use treatment programs with strong validity and reliability
	Create detailed, written manuals for all treatment tasks including a cuing hierarchy
	Have experts in specific treatment protocols serve as consultants as needed
	Provide ongoing professional development support
	Hold regular clinician meetings to discuss treatment questions and concerns
Monitoring of treatment delivery	Have fidelity coordinators observe clinicians' treatment sessions regularly (either live or video-recorded) and provide feedback
	Have clinicians complete self-assessments and peer assessments for additional feedback
Monitoring of treatment receipt	Have fidelity coordinators observe clinicians' treatment sessions regularly (either live or video-recorded) and provide feedback
	Ask participants to complete a satisfaction survey including questions about treatment delivery/receipt

(Borrelli, 2011; Gearing et al., 2011; Hinckley & Douglas, 2013). Investigators in this study have created a participant satisfaction survey that addresses some fidelity issues. Future assessment and treatment observations will also include peer- and self-assessment evaluations.

Finally, at this time, it is not clear which of the described strategies are most important in ensuring recruitment/retention and fidelity. Although some strategies seem to be helpful (e.g., changing our recruitment tactics to increase the number of appropriate referrals), other strategies may not be as necessary or need to be as frequent (e.g., continuing to observe so many SLP assessment and treatment sessions when compliance was relatively high from the beginning). A future review of the effect of each strategy compared to its research cost would be helpful to determine which strategies to prioritize based on the resources available. Importantly, we do not know how our recruitment and retention strategies, which according to best practices seek to develop a study identity and a sense of belonging, impact psychosocial functioning and consequently what impact that improved psychosocial status might have on treatment outcomes. This clinical population consistently reports social isolation following aphasia onset, and the many visits and activities associated with study membership and lab affiliation certainly provide more opportunities to ameliorate that isolation to a degree. This should be monitored and addressed in future studies.

Creating and reporting on assessment and treatment fidelity procedures and results increase confidence in research findings and makes them more applicable to evidence-based practice. The purpose of this study was to describe how investigators of a large, multifacility, aphasia intervention study planned, implemented, and continue to modify an assessment and treatment fidelity plan. General fidelity strategies that were implemented in this study are listed in Tables 2, 3, and 4. To assure ongoing consistency in assessment and treatment fidelity, future and specific goals for this clinical research are (a) to provide ongoing support and training for clinicians and raters as needed; (b) to expand our professional referral network and community outreach activities for quality participant referrals and to continue to accommodate and provide resources for appropriate, motivated participants for retention; and (c) to continue to collect and manage data in a consistent manner and within a secure system.

Acknowledgments

This study was funded by National Institute on Deafness and Other Communication Disorders Grants P50 DC014664 (awarded to PI: Julius Fridriksson) and T32 DC014435 (Trainee: Alexandra Basilakos), and National Institute of General Medical Sciences Grant P20 GM109089 (Mentored Investigator: Jessica D. Richardson). We are grateful to the speech-language pathologists who dedicate their time and expertise to this study: Allison Croxton, Anna Doyle, Michele Martin, Katie Murphy, and Sara Sayers. We would also like to acknowledge the graduate research assistants in the University of South Carolina Master's of Speech-Language Pathology program who serve as raters for our outcome measures.

References

Aphasia Lab-USC. (n.d.). *Home* [Facebook page]. https://www.facebook.com/StrokeRecovery Project/

Aphasia Lab at the University of South Carolina. (n.d.). [Home page]. https://web.asph.sc.edu/aphasia/

Bak, T. H., & Hodges, J. R. (2003). Kissing and dancing—A test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation: Preliminary results in patients with frontotemporal dementia. *Journal of Neurolinguistics, 16*(2–3), 169–181.

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.

Barthel, G., Meinzer, M., Djundja, D., & Rockstroh, B. (2008). Intensive language therapy in chronic aphasia: Which aspects contribute most? *Aphasiology, 22*(4), 408–421.

Basso, A. (2005). How intensive/prolonged should an intensive/prolonged treatment be? *Aphasiology, 19*(10–11), 975–984.

Behn, N., Hilari, K., Marshall, J., Simpson, A., Northcott, S., Thomas, S., Flood, C., Goldsmith, K., & McVicker, S. (2018). SUpporting well-being through PEeR-Befriending (SUPERB) trial: An exploration of fidelity in peer-befriending for people with aphasia. *Aphasiology, 32*(Suppl. 1), 21–23.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (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.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*(4), 987–993.

Borrelli, B. (2011). The assessment, monitoring and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry, 71*(S1), S52–S63.

Bothe, A. K., & Richardson, J. D. (2011). Statistical, practical, clinical, and personal significance: Definitions and applications in speech-language pathology. *American Journal of Speech-Language Pathology, 20*(3), 233–242.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 2016*(6), CD000425.

Breitenstein, S., Robbins, L., & Cowell, J. M. (2012). Attention to fidelity: Why is it important? *Journal of School Nursing, 28*(6), 407–408.

Brogan, E., Ciccone, N., & Godecke, E. (2019). Treatment fidelity in aphasia randomised controlled trials. *Aphasiology, 33*(7), 759–779. https://doi.org/10.1080/02687038.2019.1576442

Center for the Study of Aphasia Recovery. (n.d.). [Home page]. https://cstar.sc.edu/

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of aphasia therapy: Evidence and efficacy. *Current Neurology and Neuroscience Reports, 11*(6), 560–569.

Cherney, L. R., Patterson, J. P., Raymer, A. M., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research, 51*(5), 1282–1299.

Claridge, A. M. (2014). Efficacy of systemically oriented psychotherapies in the treatment of perinatal depression: A meta-analysis. *Archives of Women's Mental Health, 17*(1), 3–15.

Collaborative Institutional Training Initiative (CITI) Program. (n.d.). *Research ethics and compliance training*. Retrieved from https://about.citiprogram.org/en/homepage/

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490–520.

Doyle, P., Hula, W., Hula, S., Ross, K., Wambaugh, J., Schumacher, J., & Roehrig, A. (2012, May). *The Aphasia Communication Outcome Measure: Item reduction, scaling, and concurrent validity of self-reported communicative functioning in aphasia*. Paper presented at the Clinical Aphasia Conference, Lake Tahoe, CA, United States.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350.

Elsner, B., Kugler, J., Pohl, M., & Mehrholz, J. (2012). Transcranial direct current stimulation (tDCS) for improving aphasia in patients with aphasia after stroke. *Cochrane Database of Systematic Reviews, 2019*(5), 462–493. CD009760. https://doi.org/10.1002/14651858.CD009760.pub3

Fridriksson, J., Richardson, J. D., Fillmore, P., & Cai, B. (2012). Left hemisphere plasticity and aphasia recovery. *NeuroImage, 60*(2), 854–863.

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79–88.

Girolametto, L., Weitzman, E., & Greenberg, J. (2012). Facilitating emergent literacy: Efficacy of a model that partners speech-language pathologists and educators. *American Journal of Speech-Language Pathology, 21*(1), 47–63.

Godecke, E., Armstrong, E., Middleton, S., Rai, T., Cadilhac, D., Ciccone, N., Whitworth, A., Rose, M., Hankey, G. J., Holland, A., & Benrhardt, J. (2015). Assessment of therapy fidelity processes in the Very Early Rehabilitation in SpEech (VERSE) clinical trial. *International Journal of Stroke, 10*(S2), 424–425.

Gresham, F. M., Gansle, K. A., Noell, G. H., & Cohen, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review, 22*(2), 254–272.

Gul, R. B., & Ali, P. A. (2010). Clinical trials: The challenge of recruitment and retention of participants. *Journal of Clinical Nursing, 19*(1–2), 227–233.

Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the adolescent alcohol prevention trial. *Journal of Studies on Alcohol and Drugs, 52*(6), 568–579.

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), S279–S284.

Holland, A. L., Fromm, D. S., DeRuyter, F., & Stein, M. (1996). Treatment efficacy: Aphasia. *Journal of Speech and Hearing Research, 39*(5), S27–S36.

Howard, D., & Patterson, K. E. (1992). *The Pyramids and Palm Trees Test: A test of semantic access from words and pictures*. Thames Valley Test Company.

Hróbjartsson, A., Thomsen, A. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaud, P., & Brorson, S. (2013). Observer bias in randomized clinical trials with measurement scale outcomes: A systematic review of trials with both blinded and nonblinded assessors. *Canadian Medical Association Journal, 185*(4), E201–E211.

Kay, J., Lesser, R., & Coltheart, M. (2009). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia*. Psychology Press.

Kelly, J. A., Heckman, T. G., Stevenson, L. Y., Williams, P. N., Ertl, T. Hays, R. B., Leonard, N. R., O'Donnell, L., Terry, M. A., Sogolow, E. D., & Neumanns, M. S. (2000). Transfer of research-based HIV prevention interventions to community service providers: Fidelity and adaptation. *AIDS Education and Prevention, 12*(Suppl. 5), 87–98.

Kertesz, A. (2007). *The Western Aphasia Battery–Revised*. Pearson.

Kertesz, A., & McCabe, P. (1977). Recovery patterns and prognosis in aphasia. *Brain, 100*(1), 1–18.

Kladouchou, V., Papathanasiou, 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 Communication Disorders, 52*(6), 733–749.

Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology, 9*(1), 19–43.

Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery, & Psychiatry, 79*(5), 530–534.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Erlbaum.

Martin, N., Kohen, F. P., & Kalinyak-Fliszar M. (2010). *A processing approach to the assessment of language and verbal short-term*

memory abilities in aphasia. Paper presented at the Annual Clinical Aphasiology Conference, Charleston, SC, United States.

Maxfield, L., & Hyer, L. (2002). The relationship between efficacy and methodology in studies investigating EMDR treatment of PTSD. *Journal of Clinical Psychology, 58*(1), 23–41.

McDonald, P. W. (1999). Population-based recruitment for quit-smoking programs: An analytic review of communication variables. *Preventive Medicine, 28*(6), 545–557.

Milburn, T. F., Hipfner-Boucher, K., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Effects of coaching on educators' and preschoolers' use of references to print and phonological awareness during a small-group craft/writing activity. *Language, Speech, and Hearing Services in Schools, 46*(2), 94–111.

Miller, F. G., Rosenstein, D. L., & DeRenzo, E. G. (1998). Professional integrity in clinical research. *Journal of the American Medical Association, 280*(16), 1449–1454.

Moss Rehabilitation Research Institute. (2016). *Philadelphia Naming Test (PNT)*. https://mrri.org/philadelphia-naming-test/

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315–340.

Nigg, C. R., Allegrante, J. P., & Ory, M. (2002). Theory-comparison and multiple-behavior research: Common themes advancing health behavior research. *Health Education Research, 17*(5), 670–679.

Patel, M. X., Doku, V., & Tennakoon, L. (2003). Challenges in recruitment of research participants. *Advances in Psychiatric Treatment, 9*(3), 229–238.

Pereletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of studies and examination of associated factors. *Journal of Consulting and Clinical Psychology, 75*(6), 829–841.

Plowman, E., Hentz, B., & Ellis, C., Jr. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice, 18*(3), 689–694.

Richardson, J. D., Hudspeth Dalton, S. G., Shafer, J., & Patterson, J. (2016). Assessment fidelity in aphasia research. *American Journal of Speech-Language Pathology, 25*(4S), S788–S797.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24,* 121–133.

Robey, R. R. (1994). The efficacy of treatment for aphasic persons: A meta-analysis. *Brain and Language, 47*(4), 582–608.

Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research, 41*(1), 172–187.

Rojavin, M. A. (2005). Recruitment index as a measure of patient recruitment activity in clinical trials. *Contemporary Clinical Trials, 26*(5), 552–556.

Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competencies required to deliver effective cognitive and behavioral therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy, 36*(2), 129–147.

Salis, C., Hwang, F., Howard, D., & Lallini, N. (2017). Short-term and working memory treatments for improving sentence comprehension in aphasia: A review and a replication study. *Seminars in Speech and Language, 38*(1), 29–39.

Sánchez, V., Steckler, A., Nitirat, P., Hallfors, D., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Education Research, 22*(1), 95–107.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton-Mifflin.

Solomon, D., Battistich, V., Watson, M., Schaps, E., & Lewis, C. (2000). A six-district study of educational change: Direct and mediated effects of the child development project. *Social Psychology of Education, 4*(1), 3–51.

Stitt, J. K., Simonds, C. J., & Hunt, S. K. (2003). Evaluation fidelity: An examination of criterion-based assessment and rater training in the speech communication classroom. *Communication Studies, 54*(3), 341–353.

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51,* 43–50.

Szaflarski, J. P., Allendorfer, J. B., Banks, C., Vannest, J., & Holland, S. K. (2013). Recovered vs. not-recovered from post-stroke aphasia: The contributions from the dominant and non-dominant hemispheres. *Restorative Neurology and Neuroscience, 31*(4), 347–360.

Thompson, C. K. (2012). *Northwestern Assessment of Verbs and Sentences (NAVS)*. Northwestern University.

Volkmer, A., Spector, A., Warren, J. D., & Beeke, S. (2018). The 'Better Conversations with Primary Progressive Aphasia (BCPPA)' program for people with PPA (Primary Progressive Aphasia): Protocol for a randomised controlled pilot study. *Pilot and Feasibility Studies, 4*(1), 158.

Walker, G. M., & Schwartz, M. F. (2012). Short-form Philadelphia Naming Test: Rationale and empirical evaluation. *American Journal of Speech-Language Pathology, 21*(2), S140–S153.

Web Data Coordination Unit. (n.d.). https://dcu.musc.edu/

Wechsler, D. (2008). *Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV)*. NCS Pearson.

Worrall, L., Ryan, B., Hudson, K., Kneebone, I., Simmons-Mackie, N., Khan, A., Hoffman, T., Power, E., Togher, L., & Rose, M. (2016). Reducing the psychosocial impact of aphasia on mood and quality of life in people with aphasia and the impact of caregiving in family members through the Aphasia Action Success Knowledge (Aphasia ASK) program: Study protocol for a randomized controlled trial. *Trials, 17*(1), 153.

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Deborah, H., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology, 25*(3), 309–322.