The assessment of fidelity in a motor speech-treatment approach

Deborah Hayden[1], Aravind Kumar Namasivayam[2], Roslyn Ward[3,4]

[1]The PROMPT Institute, Santa Fe, NM, USA, [2]Department of Speech-Language Pathology, University of Toronto, Canada, [3]Department of Paediatric Rehabilitation, Princess Margaret Hospital for Children, WA, Australia, [4]Faculty of Medicine, School of Paediatrics and Child Health, Dentistry and Health Sciences, The University of Western Australia, Perth, WA, Australia

Objective: To demonstrate the application of the constructs of treatment fidelity for research and clinical practice for motor speech disorders, using the Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT) Fidelity Measure (PFM). Treatment fidelity refers to a set of procedures used to monitor and improve the validity and reliability of behavioral intervention. While the concept of treatment fidelity has been emphasized in medical and allied health sciences, documentation of procedures for the systematic evaluation of treatment fidelity in Speech-Language Pathology is sparse.

Methods: The development and iterative process to improve the PFM, is discussed. Further, the PFM is evaluated against recommended measurement strategies documented in the literature. This includes evaluating the appropriateness of goals and objectives; and the training of speech–language pathologists, using direct and indirect procedures. Three expert raters scored the PFM to examine inter-rater reliability.

Results: Three raters, blinded to each other's scores, completed fidelity ratings on three separate occasions. Inter-rater reliability, using Krippendorff's Alpha, was >80% for the PFM on the final scoring occasion. This indicates strong inter-rater reliability.

Conclusion: The development of fidelity measures for the training of service providers and treatment delivery is important in specialized treatment approaches where certain 'active ingredients' (e.g. specific treatment targets and therapeutic techniques) must be present in order for treatment to be effective. The PFM reflects evidence-based practice by integrating treatment delivery and clinical skill as a single quantifiable metric. PFM enables researchers and clinicians to objectively measure treatment outcomes within the PROMPT approach.

Keywords: PROMPT, Fidelity, Treatment, Motor speech disorders, Integrity, Evidenced-based practice, Intervention, Inter-rater reliability

Introduction

The implementation of evidence-based practice (EBP) principles in the identification and selection of the most appropriate treatment for a given client is well established in the field of speech–language pathology (Dollaghan, 2008) and requires a speech–language pathologist (SLP) to critically evaluate the best available scientific evidence. Traditionally, this evaluation has been formulated on the basis of methodological rigor and strength of findings associated with the dependent variable (outcome measures). These strengths may include reliability of the outcome measures, control of subjective bias, and hierarchy of evidence (Dollaghan, 2008). While these elements are essential to understanding threats to internal validity, they do not address the quality of the independent variable, that is, the fidelity of the intervention reportedly administered (Schlosser, 2002; Kaderavek and Justice, 2010).

What is treatment fidelity?

Treatment fidelity refers to the methodological strategies utilized to monitor the reliability and validity of therapy interventions. It is inextricably linked to the framework of EBP and defined as '… the degree to which administration of a treatment corresponds to the prototype treatment, also referred to as the "gold standard" implementation' (Kaderavek and Justice, 2010, p. 369). The underlying assumption is that the best possible outcomes for a client can only be achieved when an empirically supported treatment is delivered in a systematic manner with high fidelity (Kaderavek and Justice, 2010).

Correspondence to: Aravind Kumar Namasivayam, Department of Speech-Language Pathology, 160-500 University Avenue, Rehab Sciences Bldg., Toronto, ON M5G1V7, Canada.
Email: a.namasivayam@utoronto.ca

While treatment fidelity has been referred to using a variety of terms (e.g. treatment integrity, procedural integrity, treatment quality, procedural fidelity), this paper will differentiate between treatment quality and procedural fidelity (Kaderavek and Justice, 2010). Treatment quality is the skillfulness with which the clinician delivers a given treatment; that is, the clinician's ability to adjust or customize certain active ingredients of an intervention according to the needs of a client (Mihalic, 2004). For example, a clinician may increase the amount of speech motor practice along with auditory/visual cues to support the acquisition of sound sequences in a child with apraxia of speech relative to a child who has a speech sound disorder without verbal apraxia. Procedural fidelity, on the other hand, refers to the clinician's adherence to the prescribed intervention procedures and techniques. Both these aspects of treatment fidelity require assessment as even an excellent intervention with strong empirical support for its efficacy and effectiveness may not yield expected outcomes if the intervention is not delivered with high fidelity (Kaderavek and Justice, 2010).

The importance of treatment fidelity

The goal of treatment research is to assess the causal relationship between the intervention administered and the outcome measures. The quantity and quality of treatment administered has been shown to strongly correlate with effect sizes and influences the validity and power of intervention studies (e.g. Otterloo *et al.*, 2006). Furthermore, systematic intervention delivered with high fidelity has been shown to result in better and more consistent outcomes (e.g. Günther and Hautvast, 2010; Schlosser, 2002). The failure to report treatment fidelity negatively impacts upon our ability to evaluate the efficacy of an intervention under investigation. As stated by Borrelli *et al.* (2005), 'the cost of inadequate fidelity can be rejection of powerful treatment programs or acceptance of ineffective programs' (p. 852), as one is unable to determine whether the lack of treatment effect is due to deficits inherent in the treatment program or excessive alteration from the 'gold standard.' Given the above, reporting treatment fidelity is essential to the interpretation of treatment outcomes.

The establishment of treatment fidelity

The literature indicates key mutually exclusive components essential to the establishment of treatment fidelity (Schlosser, 2002; Borrelli *et al.*, 2005; Bellg *et al.*, 2004; Whyte and Hart, 2003; Kaderavek and Justice, 2010). These include explicit, verifiable and theoretically grounded treatment protocols (e.g. treatment manual), adequate training and supervision in the implementation of the treatment protocol (e.g.

supervision and certification) and systematic demonstration of adherence to the treatment protocol. Borrelli *et al.*, (2005) assessed the reporting of fidelity across 10 years of health behavior research, between 1990 and 2000. Of 342 articles evaluated, as few as 35% reported use of a treatment manual, 22% provided supervision and 27% checked adherence to the delivery of the intervention protocol. Currently, the gold standard in assessing the delivery of an intervention protocol is the administration of an evaluation protocol/checklist by a trained and reliable coder, according to *a priori* criteria (Bellg *et al.*, 2004; Schlosser, 2002) and blinded to the intervention.

One treatment approach that has demonstrated the implementation of these key strategies for establishing and measuring treatment fidelity is Prompts for Restructuring Oral Muscular Phonetic Targets (i.e. the PROMPT approach). PROMPT is a motor-speech treatment approach framed within the principles of Dynamic Systems Theory (Thelen, 2005) that strives to achieve normalized speech movement patterns via hierarchical goal selection and the use of coordinated multi-sensory inputs during task-related production of contextual and age-appropriate lexicon. PROMPT not only addresses speech production, but the development and organization of speech motor behavior as a coordinated action across several interrelated domains *viz*., physical-sensory, cognitive-linguistic, and social-emotional (Hayden *et al.*, 2010). The approach was developed by Chumpelik-Hayden (1984) and first reported as a single case study.

Treatment fidelity within the PROMPT approach

The PROMPT approach has implemented three strategies recommended in the literature for the establishment and measurement of intervention fidelity. These include manualization of the intervention, consistency in training and mentoring; and the development of a fidelity assessment tool with *a priori* criteria for the evaluation of fidelity in treatment delivery (i.e. the PROMPT Fidelity Measure).

To date, three peer-reviewed studies (Rogers, *et al.*, 2006; Dale and Hayden, 2013; Ward *et al.*, 2013), examining the effectiveness of PROMPT intervention, have reported treatment fidelity consistent with guidelines recommended in the literature (Schlosser, 2002; Borrelli *et al.*, 2005; Borrelli, 2011). A summary of the fidelity strategies reported in each of these studies, based on the five-part treatment fidelity framework, developed by the National Institutes of Health's Behavioral Change Consortium (Borrelli *et al.*, 2005; Borrelli, 2011), is provided in Table 1. This table illustrates that all three studies met at least three of the five fidelity strategies recommended by Borrelli *et al.* (2005; Borrelli, 2011). That is, they all report

Table 1 Components of treatment fidelity in three treatment studies evaluating the PROMPT approach

Fidelity components	Rogers *et al*. (2006)	Dale and Hayden (2013)	Ward *et al*. (2013)
Study design	Single subject research design	Single subject research design	Single subject research design
Study population	Nonverbal toddlers and preschoolers with autism	Preschool children with childhood apraxia of speech	Children with cerebral palsy
Number of participants	10	4	6
Treatment design			
Provide information about treatment duration and dosage:			
Length of contact (minutes)	60 minutes	50 minutes	45
Number of contacts	12	16	20
Content of treatment	PROMPT Intervention	PROMPT – full	PROMPT Intervention as described within the intervention manual
Duration of contact over time	12 weeks	8 weeks	x2 blocks, 10 weeks each
Provide information about treatment dose in comparison condition:		PROMPT – without tactile input	
Length of contact (minutes)	60 minutes	50 minutes	N/A
Number of contacts	12	16	N/A
Content of treatment	DENVER model	PROMPT – without tactile input	N/A
Duration of contact over time	12 weeks	8 weeks	N/A
Mention of provider credentials	Yes	Yes	Yes
Mention of a theoretical model	Yes. Dynamic Systems Theory	Yes. Dynamic Systems Theory	Yes. Dynamic Systems Theory
Training providers			
Description of how providers were trained	Treatment developers viewed and coded tapes of the therapist (frequency not stated), visited the site quarterly, and provided telephone supervision monthly	Not provided	All therapists had completed: Introduction to Technique training, used the technique for a minimum of 9 months, completed a case study requiring assessment by a PROMPT Instructor, and attended a mentoring day held by the developer of the technique. Three therapists also completed PROMPT Bridging to Intervention
Standardized provider training	Not specified	Not specified. PROMPT certification requires completion of the Introduction to Technique and Bridging Technique to Intervention workshops, as well as completion of a 4-month certification project	Yes
Measured provider skill post training	Three consecutive fidelity measures at 85% of greater required before commencing the intervention	Two certified SLPs with extensive experience using PROMPT	Four SLPs administered the intervention – three were trained to PROMPT Bridging to Intervention level, one was trained to PROMPT Introduction to Technique level. All therapists met 80% fidelity prior to commencing the intervention
Described how provider skills were maintained over time	Single blinded assessor completed the PFM on 25% of therapist sessions. Fidelity of at least 85% was maintained	Single blinded assessor completed the PFM two occasions during the intervention. Fidelity above 95% was maintained	Single blinded assessor completed the PFM. Two fidelity measures per participant, per intervention phase were taken to generate a total of four ratings per participant. Fidelity ranged between 77.7% and 97%
Delivery of treatment			
Method to ensure that the content of the intervention is delivered as specified	Administration of the PFM	Administration of the PFM	Administration of the PFM

Continued

Table 1 *Continued*

Fidelity components	Rogers *et al.* (2006)	Dale and Hayden (2013)	Ward *et al.* (2013)
Method to ensure that the dose of the intervention is delivered as specified	Not specified	Not specified	Not specified
Mechanism to assess provider adhered to the intervention plan	PFM scored based on video-recording of randomly selected intervention sessions	PFM scored based on video-recording of randomly selected intervention sessions	PFM scored based on video-recording of randomly selected intervention sessions
Assess nonspecific treatment effects	Not specified	Not specified	Not specified
Use of treatment manual	Yes	Yes	Yes
Receipt of treatment			
Assessed subject comprehension of the intervention during the intervention period	PFM	PFM	PFM
Included a strategy to improve subject comprehension of the intervention	Not specified	Not specified	Not specified
The participants' ability to perform the intervention skills will be assessed during the intervention	Outcome measures stated and data analyses of these measures provided	Outcome measures stated and data analyses of these measures provided	Outcome measures stated and data analyses of these measures provided
Assessed participant's ability to perform the intervention skills	Outcome measures stated and data analyses of these measures provided	Outcome measures stated and data analyses of these measures provided	Outcome measures stated and data analyses of these measures provided
Enactment of treatment skills			
Assessed subject performance	Reporting of outcome measures	Reporting of outcome measures	Reporting of outcome measures
Assessed strategy to improve subject performance	Not specified	Not specified	Not specified

Note. PFM, PROMPT Fidelity Measure; N/A, not applicable.

information regarding treatment fidelity strategies including training provided and the evaluation of treatment delivery using the PROMPT Fidelity Measure (PFM).

In each study fidelity to the intervention was calculated at 85% or greater, as rated by a single blinded assessor across the phases of the treatment study. However, other than mentioning that a Likert style rating was used, none of these three studies provide information on what the items were, how the scores were weighted, what domains and dimensions were assessed, how the final scores were calculated or the inter-reliability of the fidelity measure itself.

Given that the evaluation of treatment fidelity is dependent on the psychometric soundness of the fidelity instrument, the purpose of this paper is to document progress toward establishing the psychometric properties of the PFM. Specifically, the question addressed in this paper is: what is the inter-rater reliability of the PFM? Reliability of the PFM was assessed through measures of inter-rater agreement.

Method
Participants
Participants consisted of three raters: the first and third author and another independent rater. All participants were certified SLPs specializing in developmental motor speech disorders with more than ten years' experience in using the PROMPT approach.

The PROMPT Fidelity Measure
The PMF (see Appendix A) consists of 36 items and utilizes a 4-point Likert style rating system based on behavioural frequencies, where a score of 1 indicates that a behaviour is rarely observed while 4 indicates that a behaviour is always observed. Thus, a clinician is judged based on his/her ability to intentionally use a therapeutic strategy that results in an observable change in the client's behavior. The PMF represents each of the domains (physical-sensory, cognitive-linguistic, and social-emotional) described in the PROMPT conceptual framework and is a composite view of a clinician's adherence to nine core elements of the PROMPT protocol and philosophy (Hayden *et al.*, 2010). Examples of some of the key elements within each domain and the total possible points for each domain on the PFM include:

Physical-sensory (e.g. whether appropriate prompting is given at the right time and for the right purpose). Total of 60 points possible.

Cognitive-linguistic (e.g. activities that are used are at the appropriate cognitive level to engage the child; while language used matches/slightly exceeds the

receptive language of the child). Total of 32 points possible.

Social-emotional (e.g. clinician interaction optimizes child arousal & joint attention; clinician consistently reinforces positive behavior). Total of 36 points possible.

Therapy set up and strategies (e.g. work space is used appropriately given the nature of the activity). Total of 16 points possible.

To compute the total fidelity score, all items are tallied and converted to a percentage score. Clinicians must earn a minimum of 100 points out of 144 total points (~70%) to pass PROMPT fidelity (certification) requirements. Clinicians may re-do selected sections if they achieved a score between 80 and 99 (~55–69%). A score of ≤79 (<55%) requires a complete re-do of the treatment project (a new peer reviewer is assigned).

The PFM is a direct measure completed either live or via video recording. Live assessments are not generally carried out due to the time and personnel costs involved. Typically, the SLP being assessed for fidelity is required to videotape four sessions across the entire treatment block (e.g. one videotaping session every 2 weeks in an 8-week treatment block) and uploads the videos on to a secure server at the PROMPT Institute. Approximately 20% of these videos are randomly chosen and rated for fidelity by any PROMPT Instructor from any geographic location logging into the system. The instructors view tapes, score fidelity and provide feedback to the clinician, and iterate the process if fidelity scores are below 70%. To score each item, the rater assigns a rating (1–4) based on their judgment as to whether the clinician delivering the PROMPT intervention is consistently applying a certain technique or strategy.

Currently, to administer the PFM, a rater must be trained to the level of a PROMPT Instructor. This means the rater has passed the PROMPT certification process, attended the Instructor training program and attends yearly Instructor updates run by the Prompt Institute. The Instructor meetings and the certification project/process are both recognized by the Continuing Education Board of the American Speech-Language-Hearing Association and clinicians are typically allowed to accrue continuing education credits.

Procedure

Reliability of the three raters was assessed through measures of inter-rater agreement. The process was as follows:

Step 1. Operationalization of the definitions. The intent of each item of the PFM was discussed and the definitions were operationalized through a consensus approach. Once all the items within one domain (e.g. physical sensory domain) had been defined, a de-identified certification project was randomly selected and scored independently by each rater. One rater collated the fidelity measures and identified items with poor agreement. The definitions for these items were further refined. This process was repeated on three occasions (see Table 2), until all items of the PFM reached 100% consensus.

Step 2. Item testing and revision. The PFM used a frequency of behavior rating system for all items. However, during the process of operationalizing the definitions, it became apparent that consensus was improved with some items being assigned a cumulative or criterion-based score. Therefore, the 4-point Likert system, based on frequency, criterion and cumulative was implemented. The following example illustrates the process of fine-tuning the definitions and the implementation of a cumulative score format, using item 2 in the physical-sensory domain of the PFM:

> 'The child is positioned closely to the clinician for adequate prompting and support'.

Initially (consensus meeting 1), two key descriptors were identified as essential:
 a. The child/clinician pair are physically positioned in close proximity to allow the clinician a *neutral or 'at rest' shoulder* position, and
 b. The child/clinician is positioned for appropriate *head/neck alignment*.

However, subsequent to scoring a de-identified certification project, it became apparent that two additional descriptors were required. Thus, the descriptors for this item were further refined (consensus meeting two), as follows:
 a. Positioning between the clinician and client allows for appropriate *eye contact*.
 b. Positioning between the clinician and client enable *joint interaction with the materials*;
 c. The clinician is appropriately positioned to allow clinician a *neutral or 'at rest' shoulder* position;
 d. Positioning between the clinician and client is comfortable with *good head/neck alignment* for the client.

A single point has been allocated to each item to a maximum of 4 points.

Step 3. Pilot testing to test psychometric properties. Once the definitions for each item had been completed to consensus, inter-rater reliability measures were piloted by a four person team. This process involved the same three experienced PROMPT instructors and an SLP researcher familiar with psychometrics and test construction. The three raters blinded to each other's scores viewed and assessed fidelity on a randomly chosen treatment sample video that was uploaded on to the PROMPT Institute's secure server. Each rater then emailed the fidelity scores directly to the researcher

Table 2   Raw data for three raters across key items on the PFM for occasion 1, 2 and 3*

PFM Item number	Occasion 1				Occasion 2				Occasion 3			
	RR1	R2	RR3	Disagreement	RR1	R2	RR3	Disagreement	RR1	R2	RR3	Disagreement
Physical-sensory												
1	1	1	1		1	1	1		1	1	1	
2	4	4	4		4	4	4		4	4	4	
3	3	3	2	xx	3	3	3		3	3	3	
4	2	2	2		3	1	2	xx	2	2	2	
5	2	2	2		3	2	2	xx	2	2	2	
6	2	2	2		2	2	2		2	2	2	
7	3	3	3		3	2	3	xx	3	3	3	
8	4	3	4	xx	3	3	3		4	4	3	xx
9	3	4	3	xx	3	3	4	xx	4	3	4	xx
10	3	3	2	xx	3	2	3	xx	3	3	3	
11	1	1	1		1	1	1		1	1	1	
12	1	1	1		1	1	1		1	1	1	
13	1	1	1		1	1	1		1	1	1	
14	3	3	4	xx	3	4	3	xx	4	3	3	xx
15	4	4	4		2	4	4	xx	4	4	4	
Cognitive-linguistic												
1	4	4	4		4	4	4		4	4	4	
2	3	3	4	xx	3	3	3		3	3	3	
3	3	3	4	xx	3	4	3	xx	3	3	3	
4	3	3	4	xx	3	3	3		3	3	3	
5	4	4	4		3	4	4	xx	4	4	4	
6	2	2	3	xx	2	3	2	xx	2	2	2	
7	4	4	2	xx	3	4	4	xx	4	4	4	
8	1	1	3	xx	2	3	2	xx	2	1	1	xx
Social-emotional												
1	4	4	4		3	3	4	xx	4	4	4	
2	2	2	1	xx	1	1	1		3	2	2	xx
3	3	3	3		2	4	3	xxx	3	3	3	
4	4	4	3	xx	2	4	4	xx	4	4	4	
5	4	4	1	xx	3	4	4	xx	4	4	4	
6	1	1	4	xx	2	3	2	xx	1	1	1	
7	3	3	3		4	4	3	xx	4	3	3	xx
8	4	4	4		4	4	4		4	4	4	
9	4	4	2	xx	3	4	4	xx	4	4	4	
Therapy set-up and strategies												
1	4	4	4		4	4	4		4	4	4	
2	4	4	4		4	4	4		4	4	4	
3	4	4	4		4	4	4		4	4	4	
4	4	4	4		2	4	4	xx	1	4	4	xx

*The items with xxx indicate major disagreements (all three raters have different scores) and items with xx indicate partial disagreement (two raters agree and one disagrees).

for compilation, statistical analysis and interpretation. Following each occasion, the researcher then provided feedback on inter-rater agreement and items on which disagreements had occurred.

Statistical analyses

Inter-rater reliability coefficient Krippendorff's alpha was calculated using the online Reliability Calculator for Ordinal, Interval, and Ratio data (ReCal OIR; Freelon, 2013). The following guidelines are recommended for interpretation of the coefficient, where $\alpha \geq 0.800 =$ good reliability, acceptable conclusions from the data can be drawn; $0.667–0.8 =$ fair, tentative conclusions can be drawn from the data; and $<0.667 =$ poor, the data is not reliable (Hallgren, 2012).

Results

Table 2 contains raw data for three raters across key items for the first, second and third occasion and

illustrates the key items requiring refinement to the definitions and changes to the scoring criteria to achieve an acceptable level of reliability. The items with 'xxx' indicate major disagreements across all three raters (i.e. all three raters have different scores) and items with 'xx' indicate partial disagreement (two raters agree and one disagrees). The definitions and scoring approach for items with disagreements were refined by a consensus approach. The inter-rater reliability coefficients, using Krippendorff's alpha for each of the three testing occasions, were 0.69, 0.72, and 0.89, respectively. The data show good inter-rater reliability was established on the third testing occasion.

Discussion

The purpose of this study was to evaluate the inter-rater reliability of the PFM. The results yield preliminary findings regarding the psychometric adequacy of

the PFM. Over-all the results indicate good inter-rater reliability between three raters following operationalization of the definitions for each item contained within the PFM. Schlosser (2002) recommends nine key components essential to the assessment of treatment fidelity and associated consequences to internal, construct, and external validity. These nine items are shown in Table 3 as categorized according to intervention design, execution and managing threats to validity. The following discussion addresses how the PFM meets these components.

Intervention design

Define the independent variable operationally (item 1). The PROMPT clinical training program provides the operational definition for the independent variable and the procedural steps necessary to carry out the treatment in line with the core PROMPT principles (Hayden *et al.*, 2010). The PROMPT approach is manualized, the underlying theory is detailed, the tenets of the approach are well defined, the technique and prompts are described and taught in workshops, and clinicians are taught how to determine intervention objectives, write goals and activities, choose correct communication focus and target lexicon that are true to the approach.

Define procedural steps (item 2). The PFM contains detailed definitions, descriptions, and scoring information, enabling the clinician to self-monitor or seek mentoring to promote adherence to the approach. In

Table 3   PROMPT fidelity measure as compared with recommended measurement strategies for treatment fidelity (adapted from Schlosser, 2002, p. 44)

Recommended measurement strategy for treatment fidelity

Intervention design
1. Define the independent variable operationally: operational definitions must include verbal, physical, spatial and temporal parameters (e.g. time between cues, set up of room, verbal instructions / feedback, etc.).
2. Decide procedural steps: steps carried out during treatment monitored using a checklist (by an independent observer or the clinician).
Execution: how fidelity is measured and assessed
3. Determine an assessment method: (a) Direct assessment through behavioural observations (video-taped or live) or (b) Indirect assessment though self-monitoring/reporting.
4. Prepare data recording (fidelity scoring) sheets consistent with assessment methods (direct or indirect).
5. Ascertain the number of observations: observe between 20–40% of all sessions to have adequate representation of treatment process.
6. Calculate treatment fidelity using % accuracy scores along with % inter-observer agreement or reliability scores.
7. Report fidelity data: as overall fidelity, component fidelity, session fidelity etc.
Minimizing threats to validity
8. Minimize the reactivity of observations: therapist may behave differently whilst being watched or video-taped. Use of random schedule for recording sessions is recommended.
9. Minimize experimenter bias: self-reporting or indirect measures are inadequate by themselves

this manner the PROMPT fidelity measure meets requirements for items 1 and 2 (Table 3).

Execution

How fidelity is measured and assessed (items 3–5). The PROMPT approach utilizes both direct and indirect assessment procedures. All intervention sessions are video recorded in entirety with approximately 20% randomly selected for direct (live or video based) evaluation of the treatment process. As clinicians progress in the training, they are also encouraged to self-monitor and report clinical issues to the PROMPT instructor group where feedback and mentoring is offered.

Calculate treatment fidelity (item 6). The fidelity assessment is carried out systematically with Likert rating scales that have been operationally defined using *a priori* coding categories. The 36-item checklist-based rating system is supplemented by additional questions that target communication focus, motor speech hierarchy and priorities selected, co-morbid conditions that may have implications for intervention (e.g. tone issues), target lexicon (syllables, words, and phrases) chosen to embed speech motor movements. In the current PFM, items are rated on a 4-point Likert system, based on either frequency of occurrence, criterion (1 or 4) or cumulative (1, 2, 3, 4) points scored.

Report treatment fidelity (item 7). Each section or domain in the fidelity measure has a sub score (*Physical-sensory* = 60 *Cognitive-linguistic* = 32 *Social-emotional* = 36 and *Therapy set up and strategies* = 16; *Total* = 144) which allows the examination of overall fidelity and component fidelity (quality vs. procedural; pre-treatment clinician fidelity vs. treatment session fidelity) as percentage scores (i.e. score x of total 144 or score x of domain score, etc.). Furthermore, the itemized fidelity measure in Appendix A easily permits the calculation of both domain-by-domain or overall inter-rater reliability using point-by-point percentage agreement index (# agreements/(# agreements + #disagreements) × 100) or inter-rater reliability coefficients like Krippendorff's alpha which account for chance agreements between two or more raters using freely available software (e.g. ReCal OIR; Freelon, 2013).

Minimizing threats to validity

Minimizing reactivity of observations (item 8). Schlosser (2002) further recommends that a good fidelity measure should minimize threats to both internal and external validity, as these negatively impact treatment fidelity. Of principle concern is the change in clinician behaviour as a result of being watched or video-taped in a session (i.e. treatment may be less effective in untapped sessions). The PROMPT fidelity process somewhat ameliorates this by the requirement

for video-taping of all sessions, from which one or two sessions are randomly selected for assessment.

Minimize experimenter bias (item 9). Biases which may arise from indirect measures such as self-reporting may be ameliorated by recording and assessing video samples and by using two or more observers and calculating inter-rater reliability for the fidelity measure. The general procedure currently used for the assessment of fidelity of the PROMPT intervention is, as mentioned before, the assessment of video sample(s) by blinded/naive prompt instructors. This process minimizes threats to validity.

In summary, the PROMPT approach utilizes a systematic, manualized, independently peer-reviewed clinical certification process. The PROMPT fidelity measure utilizes a scoring system that integrates treatment delivery (procedural fidelity) and clinical skill (treatment quality) as a single quantifiable metric, which allows for the reporting of both domain-by-domain or overall fidelity scores and/or reliability. This allows for higher clinician reliability, uniform intervention delivery, and consistent treatment outcomes, which facilitate interpretation of internal/external validity study replication and dissemination (Kaderavek and Justice, 2010). Finally, the PROMPT treatment fidelity measure meets recommended measurement strategies reported in the literature (Schlosser, 2002).

Limitations

The evaluation of treatment fidelity is dependent on the psychometric soundness of the fidelity instrument. Despite the contributions of this study, it is exploratory in nature with a limited sample size. It is acknowledged that a larger sample size is required and efforts are in progress to collect inter-rater reliability across 40 independent raters. In addition, further evaluation of additional psychometric properties, including validity and clinical utility is required.

Conclusion

The development of fidelity measures for the training of service providers and treatment delivery is especially important in specialized treatment approaches where certain 'active ingredients' (e.g. specific treatment targets, therapeutic techniques and dosage) must be present in order for treatment to be effective. The construction of the PFM enables researchers and clinicians to objectively measure treatment outcomes and reflects EBP. In this study, apart from providing details of the fidelity measure, preliminary data on inter-rater reliability and the iterative process that followed to improve on the reliability scores are provided to allow future researchers to plan and develop fidelity measures for other intervention approaches. The next steps are already underway in the development of this PFM, which entails the measurement of content

validity and scaling up the study generalizability by using a larger sample size ($N = 40$) for the calculation of inter-rater reliability.

References

Bellg A.J., Borrelli B., Resnick B., Hecht J., Minicucci D.S., Ory M., *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. doi: 10.1037/0278-6133.23.5.443.

Borrelli B. 2011. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71(Suppl. 1): 52–63.

Borrelli B., Sepinwall D., Ernst D., Bellg A.J., Czajkowski S., Breger R., *et al.* 2005. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, 73(5): 852–860.

Chumpelik-Hayden D.A. 1984. The PROMPT system of therapy: theoretical framework and applications for developmental apraxia of speech. *Seminars in Speech and Language*, 5: 139–156.

Dale P., Hayden D. 2013. Treating speech subsystems in childhood apraxia of speech with tactual input: the PROMPT approach. *American Journal of Speech-Language Pathology*, 22(4): 644–661.

Dollaghan C.A. 2008. *The handbook for evidence-based practice in communication disorders.* Baltimore: Paul H. Brookes Publishing.

Freelon D. 2013. ReCal OIR: ordinal, interval, and ratio intercoder reliability as a web service'. *International Journal of Internet Science*, 8(1): 10–16.

Günther T., Hautvast S. 2010. Addition of contingency management to increase home practice in young children with a speech sound disorder. *International Journal of Language and Communication Disorders*, 45(3): 345–353.

Hallgren K.A. 2012. Computing interrater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8: 23–34.

Hayden D., Eigen J., Walker A., Olsen L. 2010. PROMPT: A Tactually Grounded model. In: Williams A.L., McLeod S., McCauley R.J., (eds.) *Interventions for speech sound disorders in children*. Baltimore: Paul H. Brookes Publishing, p. 453–474.

Kaderavek J.N., Justice L.M. 2010. Fidelity: an essential component of evidence-based practice in speech-language pathology. *American Journal of Speech Language Pathology*, 19(4): 369–379.

Mihalic S. 2004. The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth*, 4(4): 83–105.

Otterloo S.G., van der Leij A., Veldkamp E. 2006. Treatment integrity in a home-based pre-reading intervention programme. *Dyslexia*, 12(3): 155–176.

Rogers S.J., Hayden D., Hepburn S., Charlifue-Smith R., Hall T., Hayes A. 2006. Teaching young nonverbal children with autism useful speech: a pilot study of the Denver Model and PROMPT interventions. *Journal of Autism and Developmental Disorders*, 36: 1007–1024.

Schlosser R.W. 2002. On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication*, 18: 36–44.

Thelen E. 2005. Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues*, 15: 255–283.

Ward R., Strauss G.R., Leitão S. 2013. Kinematic changes in jaw and lip control of children with cerebral palsy following participation in a motor-speech (PROMPT) intervention. *International Journal of Speech-Language Pathology*, 15(2): 136–155.

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 and Rehabilitation*, 82: 639–652.

Appendix

(A) Prompt fidelity score form

Clinician:	Client:	Reviewer:	Date:

1 Motor speech hierarchy stages and priorities correctly identified
2 Communication foci correctly identified
3 Treatment stage has been correctly identified (i.e. stage 1, 2, or 3)
4 Purpose of prompt correctly identified

Physical – sensory domain
1 If tone is identified as an issue on the motor speech hierarchy, it is addressed in the intervention session
2 Child is positioned closely to the clinician for adequate prompting and physical support
3 The appropriate motor level, for the child's sensory motor capacity, is chosen
4 Appropriate prompting (parameter, syllable, complex, or surface prompting) is given at the right time and for the right purpose
5 Observed prompting technique is accurate
6 Clinician provides prompting for both (1) accuracy of motor phonemes and (2) whole word/phrase approximations
7 Frequency of prompting is appropriate
8 Child appears to understand what the goal and expected response of prompting the clinician is expecting?
9 Clinician states, asks models, and provides feedback expected response if child does not automatically produce it
10 Prompting achieves the desired effect, e.g. child imitates or is able to approximate the target in a more relaxed, refined, or intelligible manner
11 Some 'motor-phoneme' practice is seen during the session
12 If 'motor-phoneme' practice is seen, the practice is appropriate to child's motor levels

Continued

Clinician:	Client:	Reviewer:	Date:

13 If 'motor-phoneme' practice is seen, the practice is appropriate to the context of the activity in which the sounds will be embedded
14 The selected PROMPT lexicon created for use in the routine or activity is consistent with identified motor-phonemes
15 Chosen PROMPT lexicons are used functionally within the environment whenever opportunities arise
Total possible points (60)

Cognitive–linguistic domain
1 Chosen activities are at the appropriate cognitive level to engage the child
2 Chosen activities facilitate interaction and reciprocal turn-taking
3 Sounds, syllables, or words are embedded within a meaningful and appropriate context in activities
4 Child arousal and joint attention optimized by choice of materials/activities
5 Clinician uses language that matches or just slightly exceeds the receptive language level of the child
6 Clinician provides labels for associations between objects, actions, and people
7 Clinician states, asks, and models expected response if child does not automatically produce it
8 If needed, clinician changes task demands of activity to re-engage child
Total possible points (32)

Social–emotional domain
1 Overall positive affect displayed by child
2 Reciprocal turn-taking is observed in most activities between the child and clinician
3 Child's behavior indicates their ability to predict session routines
4 Clinician interaction optimizes joint arousal and joint attention
5 Clinician provides opportunities for child to communicate and interact at almost every turn (every 30–60 seconds)
6 If needed, clinician reframes task to elicit a response in the child
7 Clinician consistently reinforces positive behavior
8 Clinician appropriately addresses difficult behaviors (e.g. crying, tantrums, hitting, throwing, kicking, screaming, spitting, self-injurious)
9 Clinician affect is appropriate and natural to the situation
Total possible points (36)

Therapy set-up and strategies
1 Materials are out of child's reach, if appropriate
2 Work areas are clearly and visually delineated
3 Space is used appropriately given the nature of the activity
4 Clinician provides changes in location a few times during the session (if appropriate)
Total possible points (16)