Thoroughness and Psychometrics of Fidelity Measures in Occupational and Physical Therapy: A Systematic Review

Brittany N. Hand, Amy R. Darragh, Andrew C. Persch

OBJECTIVE. This study evaluated the thoroughness and psychometric properties of fidelity measures used by or of relevance to occupational or physical therapy.

METHOD. A systematic review of the literature was completed. Assessments used to measure occupational or physical therapy intervention fidelity were evaluated for thoroughness, reliability, validity, and clinical utility.

RESULTS. Eight fidelity measures met inclusion criteria for this systematic review. Most of the measures had moderate levels of thoroughness in the coverage of key aspects of fidelity, reported adequate to excellent reliability and validity, and were highly variable in clinical utility.

CONCLUSION. Additional research is recommended to validate existing occupational or physical therapy fidelity measures and to develop novel measures for other occupational therapy and physical therapy interventions. Clinicians and researchers must place greater emphasis on the development and implementation of fidelity measures to ensure uniformity in intervention delivery and high-quality, evidence-based care.

Hand, B. N., Darragh, A. R., & Persch, A. C. (2018). Thoroughness and psychometrics of fidelity measures in occupational and physical therapy: A systematic review. *American Journal of Occupational Therapy*, 72, 7205205050. <https://doi.org/10.5014/ajot.2018.025510>

Brittany N. Hand, PhD, OTR/L, is Assistant Professor, School of Health and Rehabilitation Sciences, The Ohio State University, Columbus. At the time of the study, she was Student, Division of Occupational Therapy, The Ohio State University, Columbus; hand.58@osu.edu

Amy R. Darragh, PhD, OTR/L, is Division Director and Associate Professor, Division of Occupational Therapy, The Ohio State University, Columbus.

Andrew C. Persch, PhD, OTR/L, is Assistant Professor, Department of Occupational Therapy, Colorado State University, Fort Collins. At the time of the study, he was Assistant Professor, Division of Occupational Therapy, The Ohio State University, Columbus.

Fidelity is the faithfulness of an intervention to its underlying therapeutic principles and clinical guidelines (Parham et al., 2007). In rehabilitation professions, such as occupational and physical therapy, clinicians and researchers rely on measures of fidelity to systematically evaluate interventions and ensure that they are replicable and unique from other interventions. High fidelity is necessary for confident conclusions to be drawn about the uniformity, quality, and efficacy of the intervention in research trials and clinical practice (Nelson & Mathiowetz, 2004). Omitting key intervention elements that are hypothesized to result in therapeutic change has been documented to lead to poorer intervention quality and outcomes (Dane & Schneider, 1998). Moreover, high intervention fidelity has been linked with improved outcomes and maintenance of skills (Eames et al., 2009; Vismara et al., 2009). Taken together, these findings demonstrate that fidelity is critical to intervention delivery because it affects outcomes.

Fidelity consists of five key components: (1) adherence, (2) quality of delivery, (3) exposure, (4) participant responsiveness, and (5) program differentiation (Dane & Schneider, 1998). *Adherence* refers to the extent to which program components are delivered as intended, whereas *quality of delivery* is a subjective aspect of treatment delivery that extends beyond delivery of prescribed content. For example, aspects of quality of delivery may include clinician enthusiasm or attitude. *Exposure* refers to the number, length, or frequency of intervention sessions or the frequency with which intervention techniques are implemented. *Participant responsiveness* includes participants' judgments about the outcomes and relevance of

an intervention (Carroll et al., 2007) and is a key aspect of intervention fidelity. Finally, *program differentiation* refers to how the intervention being delivered is different and distinguishable from other interventions (Dusenbury et al., 2003).

Implications of Intervention Fidelity

Evaluating intervention fidelity has important implications for internal validity because studies that include fidelity measures are less likely to be limited by the potential confounding of extraneous variables (Nelson & Mathiowetz, 2004). For example, if significant results are found in a study that used a fidelity measure, researchers and readers can be assured that the intervention was delivered as intended (i.e., with high fidelity) and that outcomes were not likely influenced by idiosyncratic variations in intervention delivery. Fidelity also has important implications for external validity because the use of fidelity measures facilitates the replication of interventions. For example, successful replication across research trials may be enhanced and facilitated by thorough descriptions of the intervention and manualized procedures (Mowbray et al., 2003). Finally, fidelity may affect statistical power. For example, it has been noted in the literature that it would be necessary to increase the sample size 4-fold to compensate for an intervention that is delivered with 75% fidelity (Summerfelt, 2003).

The use of fidelity measures has been documented to improve treatment outcomes; as such, these measures have important applications that extend beyond research trials to clinical practice (Bond & Salyers, 2004). For example, fidelity measures perform a quality assurance function because they confirm that an intervention is replicable and consistent across clinical facilities (Parham et al., 2007). In addition, widespread initiatives to implement fidelity measures can have long-lasting effects on intervention delivery, improving intervention fidelity for 12–24 mo (McHugo et al., 2007). Moreover, training clinicians on the use of fidelity measures may assist them in appraising the validity of research on a particular intervention (Parham et al., 2007).

Best Practices in Fidelity Measure Development

Given the importance of fidelity to research trials and clinical practice, it is not surprising that researchers are being urged to develop intervention manuals and corresponding measures of fidelity (Persch & Page, 2013). Item 5 of the CONSORT (Consolidated Standards of

Reporting Trials) statement calls for describing precise details of the interventions with “sufficient details to allow replication, including how and when they were actually administered,” making fidelity assessment an essential component of randomized controlled trials (Schulz et al., 2010, p. 2). Although practices such as training and observation without the use of an objective assessment are commonly reported methods for examining intervention fidelity (i.e., Hinckley & Douglas, 2013), these approaches do not include the use of a formal quantitative measurement tool. In this study, *fidelity measures* refer to quantitative assessments of intervention fidelity. When deciding whether to implement a fidelity measure in a research trial or clinical practice, it is critical to consider psychometric properties, such as reliability and validity. When a measure demonstrates sound psychometric properties, one can be confident that the scores consistently measure the construct that they purport to measure.

When developing a fidelity measure and testing the psychometric properties, care should be taken to select methods that are based on current best practices in measure development. Best practices for various aspects of psychometric testing are summarized in Supplemental Table 1 (available online at <http://otjournal.net>; navigate to this article, and click on “Supplemental”). *Reliability* refers to the consistency of a series of measurements and is commonly assessed via interrater and internal consistency reliability (Cronbach, 1990). *Interrater reliability* refers to the ability of different raters to obtain similar scores on an observational measure (Gliner et al., 2011). Interrater reliability is established by having multiple raters complete the assessment while observing certain episodes of behavior. Next, the scores of the raters are correlated to determine the level of agreement between raters (i.e., using Pearson’s r or Spearman’s ρ). *Internal consistency reliability* describes the average correlation among items within an assessment and may be established with interitem correlations using a correlation coefficient appropriate for the level of data (e.g., Cronbach’s α , ordinal), factor analytical approaches, or item response theory methods.

It is imperative to establish *measurement validity*, which refers to evidence for the use of a measure for its intended purpose (Gliner et al., 2011). Typically, validity is established by examining and testing the content and constructs of a measure as well as the relationship of scores to an external criterion. *Content validity* refers to whether the items that make up a measure are representative of the concept that it purports to measure and should be established via literature reviews and expert surveys or Delphi approaches (Veloza et al., 2012). The construct validity of

a measure is often explored by examining the relationship of scores to other variables. It may be beneficial to show that the assessment scores are strongly correlated with another assessment or variable that should be theoretically related (convergent validity) or that they are not strongly correlated with another assessment or variable that should be theoretically unrelated (divergent validity). Construct validity may be established using the following five methods: (1) the known-groups method, in which two groups expected to differ on the measure are indeed found to differ; (2) change in scores over time, such as pre- and posttherapist training; (3) correlation matrices with other tests purported to measure the same or a disparate construct; (4) factor analytical approaches, in which factors represent constructs; and (5) item response theory methodologies, which describe the relationship between the construct, item properties, and respondent answers.

Criterion validity refers to the relationship of scores to an external criterion and can be assessed with data collected at the same time as the administration of the novel assessment (concurrent validity) or with data collected at a later time (predictive validity; Cronbach & Meehl, 1955). *Concurrent validity* is examined by correlating scores with an external criterion measured at the same time as the fidelity measure, whereas *predictive validity* should be examined by determining the relationship of scores with intervention outcomes.

Strictly speaking, clinical utility is not a psychometric property. Nonetheless, it is very important to consider. *Clinical utility* is the clinical effectiveness, value, and ease with which a measure can be incorporated into everyday work practices (Smart, 2006). In regard to clinical utility, fidelity measures should have short training and administration time; clear, concise, and thorough manuals; simple scoring and interpretation (i.e., using cut scores); and low clinician and client active involvement (Polgar et al., 2002, as cited in Smart, 2006).

In this study, we evaluate fidelity measures used by or of relevance to occupational therapy, physical therapy, or both, in light of the aforementioned best practices. This systematic review addresses the following research question: "How thorough and psychometrically sound are existing occupational or physical therapy fidelity measures?" We answered this research question by completing the following objectives: Evaluate the extent to which occupational or physical therapy fidelity measures address key aspects of fidelity, and appraise the psychometric properties (i.e., reliability, validity, clinical utility) of occupational or physical therapy fidelity measures.

Method

Search Strategy and Screening

A comprehensive search for occupational or physical therapy fidelity measures was completed using the following databases: Academic Search Complete (date varies by journal–April 2016), CINAHL (1982–April 2016), ERIC (1966–April 2016), Medline (1949–April 2016), PsycINFO (1967–April 2016), and Web of Science (date varies by journal–April 2016). The search strategy included MeSH terms and text words for *fidelity* OR *intervention fidelity* OR *quality assurance, health care* AND *rehabilitation* OR *occupational therapy* OR *physical therapy*.

Two reviewers independently screened the titles and abstracts from articles produced by the search strategy using inclusion and exclusion criteria established a priori. Reviewers met on multiple occasions to establish consensus and read the full text of articles whenever additional information was required. Consensus was typically reached by jointly reviewing the article in light of study inclusion and exclusion criteria. Reviewers extracted the names of the eligible fidelity measures and conducted additional instrument-specific searches, using the same databases as the initial search, to ensure that no relevant articles were missed. Figure 1 provides a schematic illustration of the search strategy.

Inclusion and Exclusion Criteria

Articles were included if they met the following three criteria: (1) published in a peer-reviewed journal in English; (2) described the development of an assessment that measures intervention fidelity in occupational or physical therapy or described the psychometrics of an assessment that measures intervention fidelity in occupational or physical therapy; and (3) described a fidelity measure that is, at least in part, observational in nature. Fidelity measures relying solely on patient report or therapist report were not considered for the current study because of the documented unreliability of client self-report measures and lack of convergence of tools using client, therapist, and independent raters (Prince et al., 2008). In addition, articles were excluded if they (1) were not published in English, (2) described an evaluative assessment of intervention outcome measures or client baseline status (i.e., were not intervention fidelity measures), (3) described assessment of fidelity outside of the fields of occupational or physical therapy, or (4) described a general model for approaching intervention fidelity.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) search strategy and results of the search process.

Note. OT = occupational therapy; PT = physical therapy. Figure format from “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement,” by D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman; PRISMA Group, 2009, *PLoS Medicine*, 6, e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

Data Extraction and Quality Assessment

Once a comprehensive list of eligible fidelity measures was established, descriptive information was extracted for each fidelity measure regarding the target intervention, purpose, and number of items. Presence or absence of each of the five aspects of fidelity was coded on a dichotomous scale; fidelity measures were given a score of 1 if they included items that met the definition of that aspect, as described in the introduction. Assessment of thoroughness was completed through review of fidelity measures, fidelity measure manuals, and published articles. Additionally, reliability, validity, and clinical utility were evaluated for each measure using the a priori guidelines described next.

Reliability. The types of reliability evaluated in this review include interrater reliability, which refers to consistency of scores between raters, and internal consistency reliability, which refers to consistency among items

(Gliner et al., 2011). Reliability of included assessments was evaluated using a priori cut values for reliability coefficients. Reliability coefficients (e.g., Pearson’s r , intra-class correlation coefficients, Cronbach’s α) of $\geq .80$ were rated as excellent, $.60$ – $.79$ as adequate, and $< .60$ as poor (Law, 2004; Sakzewski et al., 2007).

Validity. The content and construct validity for all measures was evaluated according to the CanChild Outcome Measures Rating Form Manual and was classified as either excellent, adequate, or poor (Law, 2004). Content validity was assessed to be excellent if a judgmental (i.e., Delphi process) or statistical method (i.e., content validity index) was used, the measure is comprehensive, and items suited to the measurement purpose are included. Content validity was assessed to be adequate if the measure demonstrates content validity but no specific judgmental or statistical method was used (i.e., content validity was determined through expert opinion),

or poor if the instrument is not comprehensive (i.e., if there are key intervention concepts missing from the measure). Construct validity was rated as excellent if more than 2 well-designed studies showed that the instrument conforms to prior theoretical relationships among characteristics or people, adequate if 1 to 2 studies demonstrated confirmation of theoretical formulations, or poor if construct validation was not conducted or did not support the measure's construct validity.

Clinical Utility. Information about the availability, pricing, and clarity of instructions for each fidelity measure was used to describe the measure's feasibility (Morris et al., 2005; Sakzewski et al., 2007). Clarity of instructions was rated according to the CanChild Outcome Measures Rating Form manual as excellent if the instructions were clear, concise, and available. Instructions were rated as adequate if they were clear and concise but lacked some information or as poor if they were unavailable, unclear, or not concise (Law, 2004). In addition, information was gathered on ease of interpretation, which was rated as easy or complex according to the CanChild Outcome Measures Rating Form manual. Interpretation was rated as easy if it could be completed in <1 hr with a minimal amount of training and was easy for the average service provider to complete (Law, 2004). Information regarding the level of clinician and client involvement for the completion of each fidelity measure was also recorded. This information was used descriptively to summarize the acceptability of a fidelity measure (Morris et al., 2005).

Results

Description of Included Measures

The search strategy yielded a combined total of 821 articles (see Figure 1). After removing duplicates, 668 articles remained. Title and abstract review resulted in the exclusion of all but 45 articles. Full-text reviews of these articles reduced this number to 12 articles. The names of eight eligible fidelity measures were extracted from these studies. After performing additional instrument-specific searches of the literature using the fidelity measure name as the search term, 18 additional articles were identified that described the development or psychometric validation of the included measures. These eight fidelity measures met the predefined eligibility criteria:

1. Assessment of Community Integration Program Attributes (ACIPA)
2. Ayres Sensory Integration® Fidelity Measure (ASIFM)
3. Diversified Placement Approach (DPA) fidelity scale
4. Motivational Interviewing Skill Code (MISC)

5. Motivational Interviewing Treatment Integrity scale (MITI)
6. Pediatric Rehabilitation Observational Measure of Fidelity (PROF)
7. Taxonomy of Behavior Change Techniques for Smoking Cessation (Taxonomy of BCTs for SC)
8. Tool for Measurement of Assertive Community Treatment (TMACT).

Brief descriptions of each of these measures are provided in Supplemental Table 2 (online).

Thoroughness of Addressing Fidelity Constructs

Although copies of the ACIPA and DPA fidelity scale were requested from their respective authors, they were not received. Thus, assessment of thoroughness of these fidelity measures was based only on published articles. Most fidelity measures addressed three aspects of intervention fidelity, although the specific aspects that they addressed varied (Supplemental Table 3; online). The most commonly addressed constructs were adherence to the intervention and differentiation from other interventions. Exposure and participant responsiveness were the least commonly reported aspects of fidelity, each being included in only three fidelity measures. The most thorough measure, the PROF, addressed all five fidelity constructs, whereas the least thorough, the ACIPA, addressed only one aspect, which was treatment adherence.

Reliability

Two measures with data on internal consistency reliability—the ASIFM and the quality of delivery scales of the PROF—reported coefficients in the excellent range (>0.80; Supplemental Table 4; online). However, the MITI demonstrates, on average, poor internal consistency reliability, as measured by interitem correlations. Internal consistency reliability coefficients were not reported for the DPA fidelity scale, Taxonomy of BCTs for SC, TMACT, MISC, or ACIPA.

Many measures reported interrater reliability coefficients in the excellent range, including the ASIFM, DPA fidelity scale, frequency scales of the PROF, and Taxonomy of BCTs for SC (see Supplemental Table 4). Three measures, the MITI, frequency scales of the PROF, and the MISC, reported adequate values of interrater reliability. The quality of delivery scales of the PROF were reported to have poor interrater reliability. Interrater reliability coefficients were not reported for the TMACT or ACIPA.

Validity

The ASIFM, MITI, and PROF were determined to have excellent content validity because of the use of judgmental (i.e., Delphi process, formal survey of experts) or statistical (i.e., factor analysis) methods in psychometric testing and development (Supplemental Table 5; online). Content validity of all remaining fidelity measures was rated to be adequate. Most commonly, expert panels were used to establish content validity. Construct validity was rated as adequate for all fidelity measures, with the exception of the ACIPA and the Taxonomy of BCTs for SC, which had no evidence available. Construct validity of included measures was tested by means of (1) the known-groups method (ASIFM and DPA), in which groups or facilities that should theoretically differ in scores are indeed found to differ; (2) hypothesis testing (PROF, TMACT, and MISC), in which discriminant validity of domains was tested or experts were found to have high scores as hypothesized; and (3) change in scores over time (MITI, TMACT). Two fidelity measures, the MITI and TMACT, examined predictive validity (not shown in Supplemental Table 5). Specifically, higher scores on the MITI were significant predictors of smoking cessation at 3 mo, and higher scores on the TMACT were significantly associated with decreased use of health care (Cuddeback et al., 2013; McCambridge et al., 2011).

Clinical Utility

In the context of clinical utility, feasibility refers to the availability, pricing, and clarity of instructions of a measure (Morris et al., 2005; Sakzewski et al., 2007). The feasibility of included fidelity measures varied widely (Supplemental Table 6; online). All measures are available online or have been published in articles with the exception of the ACIPA and DPA fidelity scale, which are available by email request (Koop et al., 2004; Parvaneh et al., 2015). All measures are obtainable at no cost, which increases feasibility. Of the fidelity measures with manuals or instructions available for the present review, all had either excellent or adequate clarity of instructions. Copies of the ACIPA and DPA fidelity scale were requested from their respective first authors but were not received by the time of submission of this article. Included measures were administered through naturalistic and video observation; review of audio recordings, transcripts, or procedural manuals; and staff interviews. Only three fidelity measures reported the time required to complete the assessment. The PROF can be completed in 17 min, and the ACIPA and TMACT take 1.5–2 days to complete.

All but three measures, the DPA fidelity scale, PROF, and MISC, met criteria for “easy” interpretation, indicating that the measure could be interpreted in <1 hr with a minimal amount of training. Of the fidelity measures that specified requirements for score interpretation, the requirements ranged from 4 to 40 hr of training. Five of the eight fidelity measures reported cut scores or guidelines for interpreting scores (i.e., what scores constitute high vs. low fidelity).

Acceptability reflects how useful the measure is when weighed with the clinician and client level of involvement (Morris et al., 2005). Acceptability was found to be high for the ASIFM, MISC, MITI, PROF, and Taxonomy of BCTs for SC because these measures require only the clinician to record the session and no direct client participation. The ACIPA and TMACT had the lowest acceptability because they require team surveys, staff interviews, direct observation, and client interviews.

Discussion

Evaluation of intervention fidelity is a challenge for occupational therapy and physical therapy researchers and clinicians (Hildebrand et al., 2012). However, because reliable and valid fidelity measures are a useful means of ensuring uniformity of intervention delivery within research trials and clinical practice, it is imperative that psychometrically sound fidelity measures be developed and used (Moncher & Prinz, 1991). High fidelity is necessary if conclusions are to be confidently drawn about the efficacy of an intervention and has been linked with better intervention outcomes, which is the ultimate goal of rehabilitation (Bond et al., 2000; Eames et al., 2009; Moncher & Prinz, 1991; Vismara et al., 2009). This systematic review identified eight existing observational fidelity measures for diverse occupational and physical therapy interventions across various settings; these measures cover a wide range of interventions, including pediatric sensory and physical rehabilitation, mental health, supported employment, community integration, and behavior modification.

Overall, the included fidelity measures had a moderate level of thoroughness regarding coverage of the five key aspects of fidelity. Lack of thoroughness in some measures may in part be the result of the use of different theoretical approaches to fidelity during measure development. For example, the authors of the ASIFM used a “structure and process” approach, which is an alternate conceptualization of fidelity than the one used in the present review (Parham et al., 2011). Therefore, measures developed using alternate conceptualizations of fidelity may not reflect the same

constructs that we sought to identify. One aspect of fidelity that was not widely covered is exposure, which is surprising given that intensity or amount of therapy has been linked with functional outcomes (Bode et al., 2004). Going forward, it is important for developers of fidelity measures to include recommendations regarding the frequency and duration of treatment sessions.

Included fidelity measures, on average, had adequate to excellent reliability and validity. Future research on these measures should seek to evaluate the aspects of reliability or validity that have not yet been tested for that measure. For example, studies should be conducted to test the internal consistency and interrater reliability of the TMACT and ACIPA. Without reliability testing, there may be a concern that different raters watching the same intervention session will produce different scores. Regarding validity, most fidelity measures included in this review had adequate to excellent content and construct validity. However, no literature was found indicating that construct validity testing had been completed for the ACIPA or Taxonomy of BCTs for SC. The danger of using measures without validation is that there is no evidence to suggest that the measure is accurately capturing the construct of fidelity to an identified intervention. Therefore, researchers and clinicians should proceed with caution when using these fidelity measures.

Another limitation in the included fidelity measures, with the exception of the MITI and TMACT, is the absence of predictive validity testing. The authors of the measures that did examine predictive validity found that higher fidelity scores were associated with improved patient outcomes, which lends support to the notion that higher fidelity of intervention delivery improves the quality of care (Cuddeback et al., 2013; McCambridge et al., 2011). Improved patient outcomes are one key reason for using fidelity measures; thus, it will be critical to examine the extent to which higher intervention fidelity, as measured by these fidelity measures, results in improved patient outcomes going forward.

The clinical utility of a measure is another important consideration for both researchers and clinicians (Smart, 2006). Aspects of clinical utility such as ease of use, time, training and qualifications, format, interpretation, and meaning and relevance of information obtained are key factors that clinicians consider when deciding whether to use a measure in their daily practice (Polgar et al., 2002, as cited in Smart, 2006). Therefore, it is critical that developers of fidelity measures include this information in research publications and fidelity measure manuals. In the current study, multiple aspects of clinical utility were not reported or were unclear (i.e., training time, ad-

ministration time), which may be a barrier to the widespread clinical use of these measures.

Limitations and Future Directions

The current review has some potential limitations. First, additional observational measures of occupational and physical therapy intervention fidelity not published in English may have been excluded. Second, the extent to which clinicians are using these existing observational fidelity measures to ensure uniformity in the delivery of interventions in the community remains unknown because this information was outside of the scope of this review. Third, some research on fidelity measures may remain unpublished or have been published in non-peer-reviewed literature. Specifically, many large intervention trials include generic statements that fidelity was assessed without providing a detailed description of fidelity measure development and implementation. As a result, we were not able to include such measures and methods in the current study.

Going forward, we urge researchers developing and testing novel manualized interventions to create and disseminate a fidelity measure simultaneously to ensure that the intervention is completed uniformly in research trials and clinical practice and is distinguished from other interventions. Researchers assessing fidelity in the context of intervention trials should make the methods and items that they used to assess intervention fidelity available in appendices to published research articles or in separate publications (i.e., Hildebrand et al., 2012). Without reporting this information, other researchers cannot accurately replicate the study and clinicians cannot be sure that they are implementing the intervention as designed (Nelson & Mathiowetz, 2004). Additional research is recommended to validate existing occupational and physical therapy fidelity measures and to develop novel measures for other occupational and physical interventions. Clinicians and researchers must place greater emphasis on the development and implementation of fidelity measures to ensure uniformity in intervention delivery and high-quality, evidence-based care.

Implications for Occupational Therapy Research and Practice

The results of this study have the following implications for occupational therapy research and practice:

- Development and use of fidelity measures encourage researchers to deconstruct and make explicit what is inside the “black box” of rehabilitation intervention.

- Results presented here demonstrate that fidelity measures can be developed for a wide array of occupational and physical therapy interventions.
- Perhaps one reason that more occupational and physical therapy fidelity measures were not identified is the perception that many occupational and physical therapy interventions are highly individualized and thus are not suited for use of fidelity measures. However, although fidelity measures identify the key active ingredients for an intervention, they differ from step-by-step protocols (i.e., flexor tendon repair protocols such as one in Rajappa et al., 2014) and still allow for individualization and a client-centered approach.
- The ASIFM is an excellent example of how fidelity measures can be developed and used for interventions that are highly individualized. For example, the ASIFM has one item to score for the therapist presenting a child with sensory opportunities (Parham et al., 2011). The clinician has the flexibility to decide what those sensory opportunities are and how to use therapeutic skill and rapport to engage the child in those activities. In addition, the ASIFM scores whether clinicians collaborate with the child on activity choices and whether they establish a therapeutic alliance with the child, which is often an essential ingredient in occupational therapy interventions.
- Fidelity measures that promote excellence in care, while still allowing for individualization of services as appropriate, can and should be developed and used.

Conclusion

This systematic review identified, evaluated, and summarized eight observational intervention fidelity measures used by or of relevance to occupational or physical therapy. The measures evaluated the fidelity of different interventions and spanned multiple populations and practice settings. On average, the included fidelity measures demonstrated moderate thoroughness of coverage of fidelity constructs, excellent to adequate reliability and validity, and highly variable clinical utility. Suggestions for future work on these fidelity measures and the development of novel fidelity measures were discussed. Clinicians and researchers must place greater emphasis on the development and implementation of rehabilitation fidelity measures to ensure uniformity in intervention delivery and high-quality, evidence-based care. This systematic review promotes a greater emphasis on intervention fidelity by assessing the thoroughness, reliability, validity, and clinical utility of occupational and physical therapy fidelity measures and

provides recommendations for the development of novel occupational and physical therapy fidelity measures.

Acknowledgments

We greatly appreciate the Children With Hemiparesis Arm and Hand Movement Project (CHAMP) team whose passion for fidelity inspired this review: Dr. Sharon Ramey, Dr. Stephanie Deluca, Dr. Rich Stevenson, and Dr. Jane Case-Smith. This work was not funded. Dr. Darragh and Dr. Persch are supported by the CHAMP and Baby CHAMP studies, funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grants R01HD068345 and R01HD074574). At the time of the study, Dr. Hand was supported by Grant R01HD068345. No author has any competing interests to disclose with the data presented. Consent to publish is not applicable. ▲

References

- Bode, R. K., Heinemann, A. W., Semik, P., & Mallinson, T. (2004). Relative importance of rehabilitation therapy characteristics on functional outcomes for persons with stroke. *Stroke*, 35, 2537–2542. <https://doi.org/10.1161/01.STR.0000145200.02380.a3>
- *Bond, G. R., Drake, R. E., Mueser, K. T., & Latimer, E. (2001). Assertive community treatment for people with severe mental illness. *Disease Management and Health Outcomes*, 9, 141–159. <https://doi.org/10.2165/00115677-200109030-00003>
- Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H.-W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, 2, 75–87. <https://doi.org/10.1023/A:1010153020697>
- *Bond, G. R., & Salyers, M. P. (2004). Prediction of outcome from the Dartmouth Assertive Community Treatment fidelity scale. *CNS Spectrums*, 9, 937–942. <https://doi.org/10.1017/S1092852900009792>
- *Bond, G. R., Salyers, M. P., Dincin, J., Drake, R. E., Becker, D. R., Fraser, V. V., & Haines, M. (2007). A randomized controlled trial comparing two vocational models for persons with severe mental illness. *Journal of Consulting and Clinical Psychology*, 75, 968–982. <https://doi.org/10.1037/0022-006X.75.6.968>
- *Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology*, 71, 843–861. <https://doi.org/10.1037/0022-006X.71.5.843>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40. <https://doi.org/10.1186/1748-5908-2-40>

*References marked with an asterisk indicate studies included in the systematic review.

- Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: Harper & Row.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. <https://doi.org/10.1037/h0040957>
- *Cuddeback, G. S., Morrissey, J. P., Domino, M. E., Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2013). Fidelity to recovery-oriented ACT practices and consumer outcomes. *Psychiatric Services*, 64, 318–323. <https://doi.org/10.1176/appi.ps.201200097>
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23–45. [https://doi.org/10.1016/S0272-7358\(97\)00043-3](https://doi.org/10.1016/S0272-7358(97)00043-3)
- *de Jonge, J. M., Schippers, G. M., & Schaap, C. P. D. R. (2005). The Motivational Interviewing Skill Code: Reliability and a critical appraisal. *Behavioural and Cognitive Psychotherapy*, 33, 285–298. <https://doi.org/10.1017/S1352465804001948>
- *Di Rezze, B., Law, M., Eva, K., Pollock, N., & Gorter, J. W. (2013). Development of a generic fidelity measure for rehabilitation intervention research for children with physical disabilities. *Developmental Medicine and Child Neurology*, 55, 737–744. <https://doi.org/10.1111/dmcn.12114>
- *Di Rezze, B., Law, M., Eva, K., Pollock, N., & Gorter, J. W. (2014). Therapy behaviours in paediatric rehabilitation: Essential attributes for intervention with children with physical disabilities. *Disability and Rehabilitation*, 36, 16–22. <https://doi.org/10.3109/09638288.2013.775358>
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237–256. <https://doi.org/10.1093/her/18.2.237>
- Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Jones, K., Hughes, J. C., & Bywater, T. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: Care, Health and Development*, 35, 603–612. <https://doi.org/10.1111/j.1365-2214.2009.00975.x>
- Gliner, J. A., Morgan, G. A., & Leech, N. L. (2011). *Research methods in applied settings: An integrated approach to design and analysis* (2nd ed.). Boston: Routledge.
- Hildebrand, M. W., Host, H. H., Binder, E. F., Carpenter, B., Freedland, K. E., Morrow-Howell, N., . . . Lenze, E. J. (2012). Measuring treatment fidelity in a rehabilitation intervention study. *American Journal of Physical Medicine and Rehabilitation*, 91, 715–724. <https://doi.org/10.1097/PHM.0b013e31824ad462>
- 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, S279–S284. [https://doi.org/10.1044/1058-0360\(2012/12-0092\)](https://doi.org/10.1044/1058-0360(2012/12-0092))
- *Koop, J. I., Rollins, A. L., Bond, G. R., Salyers, M. P., Dincin, J., Kinley, T., . . . Marcelle, K. (2004). Development of the DPA Fidelity Scale: Using fidelity to define an existing vocational model. *Psychiatric Rehabilitation Journal*, 28, 16–24. <https://doi.org/10.2975/28.2004.16.24>
- Law, M. (2004). *Outcome Measures Rating Form guidelines*. Hamilton, Ontario, Canada: Canchild.
- *Lord, S. P., Can, D., Yi, M., Marin, R., Dunn, C. W., Imel, Z. E., . . . Atkins, D. C. (2015). Advancing methods for reliably assessing motivational interviewing fidelity using the Motivational Interviewing Skills Code. *Journal of Substance Abuse Treatment*, 49, 50–57. <https://doi.org/10.1016/j.jsat.2014.08.005>
- *Lorencatto, F., West, R., Christopherson, C., & Michie, S. (2013). Assessing fidelity of delivery of smoking cessation behavioural support in practice. *Implementation Science*, 8, 1.
- *Lorencatto, F., West, R., & Michie, S. (2012). Specifying evidence-based behavior change techniques to aid smoking cessation in pregnancy. *Nicotine and Tobacco Research*, 14, 1019–1026. <https://doi.org/10.1093/ntr/ntr324>
- Lorencatto, F., West, R., Seymour, N., & Michie, S. (2013). Developing a method for specifying the components of behavior change interventions in practice: The example of smoking cessation. *Journal of Consulting and Clinical Psychology*, 81, 528–544. <https://doi.org/10.1037/a0032106>
- *May-Benson, T. A., Roley, S. S., Mailloux, Z., Parham, L. D., Koomar, J., Schaaf, R. C., . . . Cohn, E. (2014). Interrater reliability and discriminative validity of the structural elements of the Ayres Sensory Integration® Fidelity Measure®. *American Journal of Occupational Therapy*, 68, 506–513. <https://doi.org/10.5014/ajot.2014.010652>
- *McCambridge, J., Day, M., Thomas, B. A., & Strang, J. (2011). Fidelity to motivational interviewing and subsequent cannabis cessation among adolescents. *Addictive Behaviors*, 36, 749–754.
- McHugo, G. J., Drake, R. E., Whitley, R., Bond, G. R., Campbell, K., Rapp, C. A., . . . Finnerty, M. T. (2007). Fidelity outcomes in the National Implementing Evidence-Based Practices Project. *Psychiatric Services*, 58, 1279–1284. <https://doi.org/10.1176/ps.2007.58.10.1279>
- *Michie, S., Churchill, S., & West, R. (2011). Identifying evidence-based competences required to deliver behavioural support for smoking cessation. *Annals of Behavioral Medicine*, 41, 59–70. <https://doi.org/10.1007/s12160-010-9235-z>
- *Michie, S., Hyder, N., Walia, A., & West, R. (2011). Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. *Addictive Behaviors*, 36, 315–319. <https://doi.org/10.1016/j.addbeh.2010.11.016>
- *Miller, W. R., & Mount, K. A. (2001). A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*, 29, 457–471.
- *Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). *Manual for the Motivational Interviewing Skill Code (MISC) Version 2.1*. Retrieved from <http://casaa.unm.edu/download/misc.pdf>
- *Mitcheson, L., Bhavsar, K., & McCambridge, J. (2009). Randomized trial of training and supervision in motivational interviewing with adolescent drug treatment practitioners. *Journal of Substance Abuse Treatment*, 37, 73–78. <https://doi.org/10.1016/j.jsat.2008.11.001>

- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247–266. [https://doi.org/10.1016/0272-7358\(91\)90103-2](https://doi.org/10.1016/0272-7358(91)90103-2)
- *Monroe-DeVita, M., Teague, G. B., & Moser, L. L. (2011). The TMACT: A new tool for measuring fidelity to assertive community treatment. *Journal of the American Psychiatric Nurses Association*, 17, 17–29.
- Morris, C., Kurinczuk, J. J., & Fitzpatrick, R. (2005). Child or family assessed measures of activity performance and participation for children with cerebral palsy: A structured review. *Child: Care, Health and Development*, 31, 397–407. <https://doi.org/10.1111/j.1365-2214.2005.00519.x>
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315–340. <https://doi.org/10.1177/109821400302400303>
- *Moyers, T. B., Manuel, J. K., Wilson, P. G., Hendrickson, S. M. L., Talcott, W., & Durand, P. (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, 36, 149–162. <https://doi.org/10.1017/S1352465807004055>
- *Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the Motivational Interviewing Skills Code. *Behavioural and Cognitive Psychotherapy*, 31, 177–184. <https://doi.org/10.1017/S1352465803002054>
- *Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, 28, 19–26. <https://doi.org/10.1016/j.jsat.2004.11.001>
- *Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., & Ernst, D. (2010). *Revised global scales: Motivational Interviewing Treatment Integrity 3.1. 1 (MITI 3.1. 1)*. Unpublished manuscript, University of New Mexico, Albuquerque. Retrieved from <http://www.marrch.org/associations/4671/files/02-09-MITI-3-1-1.pdf>
- *Moyers, T. B., Miller, W. R., & Hendrickson, S. M. L. (2005). How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. *Journal of Consulting and Clinical Psychology*, 73, 590–598. <https://doi.org/10.1037/0022-006X.73.4.590>
- Nelson, D. L., & Mathiowetz, V. (2004). Randomized controlled trials to investigate occupational therapy research questions. *American Journal of Occupational Therapy*, 58, 24–34. <https://doi.org/10.5014/ajot.58.1.24>
- *Parham, L. D., Cohn, E. S., Spitzer, S., Koomar, J. A., Miller, L. J., Burke, J. P., . . . Summers, C. A. (2007). Fidelity in sensory integration intervention research. *American Journal of Occupational Therapy*, 61, 216–227. <https://doi.org/10.5014/ajot.61.2.216>
- *Parham, L. D., Roley, S. S., May-Benson, T. A., Koomar, J., Brett-Green, B., Burke, J. P., . . . Schaaf, R. C. (2011). Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration® intervention. *American Journal of Occupational Therapy*, 65, 133–142. <https://doi.org/10.5014/ajot.2011.000745>
- *Parvaneh, S., Cocks, E., Buchanan, A., & Ghahari, S. (2015). Development of a fidelity measure for community integration programmes for people with acquired brain injury. *Brain Injury*, 29, 320–328. <https://doi.org/10.3109/02699052.2014.976592>
- Persch, A. C., & Page, S. J. (2013). Protocol development, treatment fidelity, adherence to treatment, and quality control. *American Journal of Occupational Therapy*, 67, 146–153. <https://doi.org/10.5014/ajot.2013.006213>
- *Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., . . . Fisher, G. (2007). An examination of the Motivational Interviewing Treatment Integrity code. *Journal of Substance Abuse Treatment*, 32, 11–17. <https://doi.org/10.1016/j.jsat.2006.07.001>
- Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 5, 56. <https://doi.org/10.1186/1479-5868-5-56>
- Rajappa, S., Menon, P. G., Kumar, M. M., & Raj, D. G. (2014). Early active motion protocol following triple Kessler repair for flexor tendon injury. *Journal of Orthopaedic Surgery*, 22, 96–99. <https://doi.org/10.1177/230949901402200124>
- Sakzewski, L., Boyd, R., & Ziviani, J. (2007). Clinimetric properties of participation measures for 5- to 13-year-old children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, 49, 232–240. <https://doi.org/10.1111/j.1469-8749.2007.00232.x>
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). *Correspondence CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials*. Retrieved from <https://www.bmj.com/content/340/bmj.c332>
- Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care*, 18, 377–382. <https://doi.org/10.1093/intqhc/mzl034>
- Summerfelt, W. T. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science*, 7, 55–61. https://doi.org/10.1207/S1532480XADS0702_2
- Veloza, C. A., Seel, R. T., Magasi, S., Heinemann, A. W., & Romero, S. (2012). Improving measurement methods in rehabilitation: Core concepts and recommendations for scale development. *Archives of Physical Medicine and Rehabilitation*, 93(Suppl.), S154–S163. <https://doi.org/10.1016/j.apmr.2012.06.001>
- Vismara, L. A., Colombi, C., & Rogers, S. J. (2009). Can one hour per week of therapy lead to lasting changes in young children with autism? *Autism*, 13, 93–115. <https://doi.org/10.1177/1362361307098516>