



Published in final edited form as:

*J Adv Nurs*. 2010 March ; 66(3): 673–682. doi:10.1111/j.1365-2648.2009.05216.x.

## Development of a tool to assess fidelity to a psycho-educational intervention

**Mi-Kyung Song, PhD RN [Assistant Professor]**

University of North Carolina at Chapel Hill School of Nursing, USA

**Mary Beth Happ, PhD RN FAAN [Professor]**

University of Pittsburgh School of Nursing, USA

**Margarete Sandelowski, PhD RN FAAN [Boshamer Distinguished Professor]**

University of North Carolina at Chapel Hill School of Nursing, USA

### Abstract

**Aim**—This paper is a description of a method to develop and conduct a customized psycho-educational intervention fidelity assessment as part of pilot work for an efficacy study. A tool designed to assess treatment fidelity to a psycho-education intervention for patients with end-stage renal disease and their surrogate decision makers, Sharing the Patient's Illness Representations to Increase Trust, is presented as an illustration.

**Background**—Despite the specificity and idiosyncrasy of individual interventions and the call to systematically evaluate treatment fidelity, how to accomplish this goal has not been clarified. Tools to adequately measure treatment fidelity are lacking.

**Methods**—We developed the Sharing the Patient's Illness Representations to Increase Trust Treatment Fidelity Assessment tool by identifying elements that were idiosyncratic to the intervention and those that could be adapted from existing tools. The tool has four components: overall adherence to the intervention content elements; pacing of the intervention delivery; overall dyad responsiveness; and, overall quality index of intervention delivery.

**Results**—Inter-rater reliability ranged from 0.80 to 0.87 for the four components. The tool showed utility in training and monitoring, such as detecting unplanned content elements delivered and the use of proscribed communication behaviours.

**Conclusion**—Psycho-educational interventions are one of the most common types of nursing interventions worldwide. Use of fidelity assessment tools customized to the individual interventions may enhance systematic evaluation of training and monitoring treatment fidelity.

### Keywords

adherence; development; fidelity; psycho-educational intervention; tool; treatment fidelity

---

© 2009 Blackwell Publishing Ltd

Correspondence to M.-K. Song: songm@email.unc.edu.

**Author contributions** MK, MB and MS were responsible for the study conception and design; responsible for the drafting of the manuscript; and made critical revisions to the paper for important intellectual content. MK performed the data collection; performed the data analysis; provided statistical expertise; obtained funding; provided administrative, technical or material support; and supervised the study.

**Conflict of interest** No conflict of interest has been declared by the authors.

## Introduction

There is growing recognition of the importance of evaluating treatment or intervention fidelity in randomized controlled trials – that is, the extent to which an intervention was delivered as conceived and planned – to arrive at valid conclusions concerning its effectiveness in achieving the target outcomes (Moncher & Prinz 1991, Waltz *et al.* 1993, Dumas *et al.* 2001, Bellg *et al.* 2004, Hawe *et al.* 2004, Harshbarger *et al.* 2006). However, practical guidelines for and exemplars of conducting evaluation of intervention fidelity are lacking. Accordingly, our purpose in this paper is to describe the development and use of a tool to assess the treatment fidelity of SPIRIT (Sharing the Patient's Illness Representations to Increase Trust), a psycho-educational intervention designed to assist end-of-life decision-making by patients with end-stage renal disease and their designated surrogate decision makers. The SPIRIT treatment fidelity assessment tool is used as an exemplar for how to develop and conduct a customized fidelity assessment of a psycho-educational intervention.

## Background

The idea of assessing treatment fidelity in behavioural and psycho-educational intervention studies was adopted from drug trials, in which strict adherence to protocols must be ensured. However, because of the dynamic and interactive nature of behavioural and psycho-educational interventions, assessing treatment fidelity in these studies is much more challenging.

Issues remain in defining and operationalizing treatment fidelity in behavioural and psycho-educational intervention protocols. The Behaviour Change Consortium confused matters by defining treatment fidelity as 'the methodological strategies to ensure' it (Bellg *et al.* 2004, p. 443), rather than the degree or level of adherence to an intervention protocol (Dumas *et al.* 2001, (Moncher & Prinz 1991, Perepletchikova & Kazdin 2005, Santacroce *et al.* 2004, Stein *et al.* 2007). Also frequently conflated are treatment fidelity and trial integrity. Treatment fidelity is only one of many factors, such as study design and subject retention, that contribute to trial integrity, or the internal validity of an intervention study. Trial integrity is not synonymous with, but rather partly depends on, treatment fidelity. Moreover, the five components (or steps) to enhance treatment fidelity (i.e. design, training, delivery, receipt and enactment) proposed by the Behaviour Change Consortium are often erroneously interpreted by researchers to mean that treatment fidelity has been established simply by including these five steps in their studies (Radziewicz *et al.* 2009, Spillane *et al.* 2007). Researchers such as Resnick *et al.* (2005a, 2005b) have provided examples of the implementation and evaluation of treatment fidelity; for example, reviewing audio-taped intervention sessions or direct observations to correct any deviations in intervention delivery. However, several elements remain unclear, such as how to conduct these assessments, how to determine which types of assessments are appropriate for which intervention, and how to define level of adherence to and deviations from interventional protocols.

Because of their dynamic and often highly individualized nature, systematic evaluation of treatment fidelity is difficult in the case of cognitive-behavioural, psycho-educational and communication interventions that involve interactions between the interventionist and patient or client (Carroll *et al.* 2000, Santacroce *et al.* 2004). Unlike delivery of pharmacotherapy, it is much more challenging to determine the dose and therapeutic elements of these interventions. This is because, to enhance receptiveness and be potentially effective, a psycho-educational intervention may need to be tailored to patients' individual needs and preferences. Therefore, the effects of an intervention are a function of what is delivered and how it is delivered rather than simply of how many of the intervention elements are delivered (Dumas *et al.* 2001, Leventhal & Friedman 2004, Harshbarger *et al.* 2006, Hawe *et al.* 2008). Furthermore, tailoring does not mean that the interventionist may extemporize during the intervention delivery; rather,

what is standardized vs. what is customized must be clearly defined and monitored. Interventionists must be evaluated on whether they appropriately used judgment and discretion.

In addition to these difficulties, the cost and time involved in developing an evaluation tool, training raters and establishing reliability of ratings contribute to the lack of reports of the systematic evaluation of treatment fidelity (Carroll *et al.* 2000). Although researchers may benefit from using existing evaluation tools, instruments developed for one intervention may not be readily adaptable to others if the guiding theory or the structure of the intervention differs, or if what was standardized and what was tailored to individual participants were not clearly defined. That is, the evaluation of treatment fidelity must be standardized and customized to the standardization and customization in the intervention being tested. However, it remains unclear how researchers should go about accomplishing this goal.

In this paper, we evaluate existing tools for monitoring treatment fidelity or for assessing patient-therapist/clinician interactions. This evaluation was carried out to assess their applicability for monitoring the treatment fidelity of the SPIRIT intervention tested in a pilot randomized controlled trial. Psycho-educational interventions are one of the most common types of nursing interventions worldwide. Such 'talking' or counselling interventions require a great deal of social and interpersonal interaction, which contributes to the quality of the interventions but also to the difficulty in training interventionists and monitoring the quality. A tool to assess treatment fidelity may facilitate a systematic evaluation of such interventions. First, we present the SPIRIT intervention. Second, we describe the development of the treatment fidelity assessment tool and its features. Finally, we illustrate the potential utility of the treatment fidelity assessment tool we adapted from existing tools and then customized to the SPIRIT intervention.

## The study

### The SPIRIT intervention

The SPIRIT intervention is an hour-long, single session, face-to-face intervention with dyads composed of an end-stage renal patient and a surrogate decision-maker delivered by a trained interventionist in an interview format. The intervention was designed to enhance discussion with patients and their chosen surrogate decision makers regarding end-of-life care. The theoretical foundation for the intervention is the Representational Approach to Patient Education (Donovan & Ward 2001, Donovan *et al.* 2007). Consistent with the representational approach (Donovan & Ward 2001), the SPIRIT intervention is a 5-step process that includes: (i) assessing illness representations, (ii) identifying and exploring gaps in knowledge and concerns, (iii) creating conditions for conceptual change, (iv) introducing replacement information and (v) summarizing. During the representational assessment step, the interventionist encourages both patient and surrogate to describe their illness representations along the dimensions of identity, timeline, consequences, controllability and spiritual and emotional representations. The goal for all parties is to achieve an understanding of the patient's illness experience and the surrogate's experience with that patient's illness. In the second step, the interventionist identifies and explores gaps in knowledge and concerns the dyad may have regarding illness progression, life-sustaining treatment and decision-making. In the third step, the interventionist encourages the members of the dyad to share their understandings about death, dying, and end-of-life care. To create conditions for conceptual change, the interventionist assists patients to identify their threshold for unacceptable outcomes of life-sustaining treatment and explores the concerns of patient and surrogate. In the fourth step, the interventionist presents end-of-life scenarios, encourages the patient to clarify goals of care and express concerns and assists surrogates to examine their willingness to take the responsibility to act on them. In the fifth and last step, the interventionist summarizes the discussion and identifies topics for further discussion. The interventionist assesses any

additional support the dyad needs (e.g. conferring with a spiritual advisor). More detailed description of the intervention with goals for each step and the guiding theory can be found elsewhere (Song *et al.* 2009).

### Existing tools relevant to assessing treatment fidelity to the SPIRIT intervention

We reviewed existing fidelity measures to assess their utility for the evaluation of fidelity to the SPIRIT intervention, including: the Method of Assessing Treatment Delivery (Leeuw *et al.* 2008); the Family-Focused Grief Therapy treatment integrity measure (Chan *et al.* 2004); the Yale Adherence and Competence Scale (Carroll *et al.* 2000); the Roter Interaction Analysis System (RIAS) (Roter *et al.* 2000, Roter & Larson 2002); various Motivational Interviewing adherence and competence measures (Miller *et al.* 2000, Miller & Rollnick 2002, Madson & Campbell 2006) and the Fidelity of Implementation Rating System (Forgatch *et al.* 2005a, 2005b).

One problem with these tools for our purpose was the blurred line between theoretical/content and process components. Content and process fidelity should be consistent with the conceptual model of the intervention (Dumas *et al.* 2001). Most of the treatment fidelity measures we reviewed were focused on monitoring a skill set or technique, which is a process component (Moncher & Prinz 1991, Pereplechikova & Kazdin 2005). Because the structure and content of the SPIRIT intervention were closely tied to the theoretical underpinnings, in addition to the process components, we needed to assess directly the use of the principles deemed essential to the representational intervention, such as exploration of illness representations to create conditions for conceptual change.

A second problem was the relative lack of attention given to identifying the delivery of unplanned components. The existing tools were focused on assessing whether and how well the prescribed components were delivered. However, to delineate the relationships between treatment dose and outcomes, it was important to evaluate not only what and how many components of the intervention were delivered but also whether and how many unplanned components were delivered. Having a measure to identify unplanned components delivered would also enhance training, monitoring and targeted retraining for interventionists' adherence to the intervention protocol.

Third, evaluation of interventionists' pacing of delivery of interventions was not included in existing tools. Given the development and testing stage of the SPIRIT intervention, we do not know yet whether any important differences exist among the five intervention steps. Thus, all steps should be treated equally in importance and be delivered within a certain range of time set for the intervention within the clinical trial protocol. Although the Motivational Interviewing Skill Code (Miller *et al.* 2000) required calculating clinician and client talk time, talk time did not capture interventionists' pacing through the intervention steps.

Fourth, the components of intervention delivery that were standardized as opposed to customized were not specified in previous treatment fidelity assessment tools. The common assumption in existing tools was that variations occur only by interventionists. This was because the focus of existing tools, except for the Motivational Interviewing fidelity measures, was on interventionist behaviours. Variations may occur, however, by participants' needs or characteristics. Based on the Representational Approach to Patient Education, the SPIRIT intervention was designed to be highly responsive to individual participants' needs. Thus, it was important to define components that were standardized and, therefore, to be delivered consistently, and components that were to be customized and, therefore, required interventionists' flexibility, discretion and judgment. To address these issues in fidelity assessment for the SPIRIT intervention, we developed the SPIRIT Intervention Fidelity Assessment Tool.

In our efforts to develop a tool to assess fidelity to a psycho-educational intervention, we used the following definition of treatment fidelity, which has been consistently used in monitoring psycho-educational interventions: the extent of adherence and competence in the interventionist's delivery of an intervention as planned (Moncher & Prinz 1991, Santacrose *et al.* 2004, Perepletchikova & Kazdin 2005, Stein *et al.* 2007). Treatment fidelity evaluation responds to the question: 'Do interventionists deliver the intervention as intended?' Treatment fidelity includes two components: adherence (the quantity of prescribed behaviours) and competence (skillfulness in the delivery of the intervention; Carroll *et al.* 2000, Moncher & Prinz 1991). That is, interventionists' behaviours are evaluated to determine the extent to which each key component was delivered as intended and if there was any unplanned component delivered (referred to as *content fidelity*), and the extent to which effective communication skills/behaviours were used and how each intervention component was delivered (referred to as *process fidelity*; Dumas *et al.* 2001).

### The SPIRIT Intervention Fidelity Assessment Tool

**Overall adherence to intervention content elements**—To define and monitor delivery of theoretical components of the SPIRIT intervention, we first determined the content elements based on the guiding theory and the goals for each step of the intervention. The 5-step SPIRIT intervention had a total of 30 elements, with Step 1 (assessing illness representation) composed of the largest number of elements (13) because it serves as the foundation for the rest of the intervention. Eleven elements were structured for patient and surrogate each (for a total of 22 elements), six elements for both patient and surrogate as a dyad and two elements for patients only. Each element included several prescribed questions that an interventionist asked of the dyad without significant modification that would change the meaning. The interventionist was allowed to choose 1–2 questions out of 3–4 prescribed questions for each element.

To evaluate content fidelity, each prescribed question was coded as *attempted* (asked as prescribed but failed to obtain an answer from the patient or surrogate); *completed* (asked as prescribed and answered); *deviated from* (asked differently and its meaning changed as a result); or as *skipped* (none asked at all). The unit of coding for content elements was the question the interventionist asked, which was compared with the prescribed questions. For example, there were three prescribed questions that could be used by the interventionist for assessing the patient's and surrogate's representation of the patient's illness during the first step (Table 1). The interventionist was required to ask at least one of these three questions of each of the patient and surrogate dyads to complete the first part (identity) of the representational assessment. For each step of the intervention, the interventionist was required to use probing questions as needed for further exploration. If none of these questions was asked, the assessment was coded as *skipped*. An example of a *deviated* element was if the interventionist explored the illness representation of the surrogate by asking 'What can you [surrogate] tell me about kidney disease?' instead of 'Tell me about your father's kidney disease'; this was coded as *deviated* because the meaning/intention of the question differed from the prescribed one. That is, the purpose of this element was to explore a surrogate's perspectives of and experiences with her father's illness rather than her general knowledge of kidney disease. Because the current intervention was at a development and testing stage, we concluded that coding each element at this level of detail (i.e. determining correspondence to the goals of each step of the intervention) was necessary. This coding system (*attempted*, *completed*, *deviated* and *skipped*) also served to identify the delivery of unplanned components.

To evaluate the overall adherence to intervention content elements, the occurrences of *attempted*, *completed*, *deviated* and *skipped* elements were simply counted. These tallies were divided by the total number of elements (30) to compute ratios (0–1); a higher score in the *completed* category indicated higher numbers of intervention elements completed. To be

considered satisfactory, ratios were to be 0.8 or higher for *completed*, 0.2 or lower for *deviated*, and 0.1 or lower for *skipped*. Because the interventionist was required to attempt more than one prescribed question for each element, a ratio of *attempted* was not computed but captured by a ratio of *completed*. We treated each element as equal in importance at this phase of development.

**Overall adherence to process components**—The process skills were conceived as a set of communication behaviours that an interventionist utilizes in response to the dyad's needs and situations while delivering the content elements (Lauver *et al.* 2002). The literature on provider–patient communication and motivational interviewing provided guidance here. Definitions of individual process skills were largely adapted from the RIAS (Roter 2001, Roter & Larson 2002) unspecified in the tools are used to assess clinician–patient interaction/communication during medical visits and Motivational Interviewing (Miller *et al.* 2000), a directive client-centred approach for behavioural modification (e.g. smoking cessation and adherence to HIV medical regimens) and assisting clients in exploring and resolving ambivalence (Miller *et al.* 2000, Miller & Rollnick 2002). We used the intervention transcripts and audio files from pilot studies (Song *et al.* 2005, in press) to identify 20 communication behaviours that interventionists might use. The interventionist was expected to use 18 positive communication behaviours as appropriate (e.g. repeat or summarize to check or confirm for assurance), while two behaviours (i.e. missed opportunities for further exploring or showing empathy; leading/assuming) were discouraged.

The unit of coding for process skills was a meaningful expression, which could be one word or a sentence. For example, a single word, 'Okay,' can be an expression of *go-on or facilitation* (Roter 2001). A sentence – 'It must have been very hard for you to make those decisions for your father' – can be coded as *empathy*. Each unit of transcribed intervention text was assigned one code (Table 2). Interventionists were expected to use timely probing, empathy, and higher level of clarification skills, such as paraphrasing or reframing, as appropriate, rather than repeating or restating. The occurrences of process skills were counted for rating overall competence of interventionists in delivering the intervention.

**Pacing of the intervention delivery**—To pace adequately without rushing or dragging, the interventionist was required to deliver each of the five steps in certain duration of time. The recommended durations for the steps were determined from an analysis of audio-recorded intervention sessions from previous pilot studies (Song *et al.* 2005, in press). For example, the duration of Step 1 could vary depending on the time necessary to build rapport with the dyad, but it was expected to last at least 15 minutes to explore all dimensions of the dyad's representations and should not last more than 25 minutes to allow for completion of the remaining steps without rushing. The actual duration of each step was rated in comparison to the recommended duration range on a 3-point rating scale from 1 (too short) to 3 (too long). A mean score of the five ratings was used for the overall adherence to pacing.

**Overall dyad's responsiveness**—Although the participants' responsiveness to the intervention is a dimension separate from interventionist's adherence or competence, participants' acceptance of an intervention may influence fidelity (Perepletchikova & Kazdin 2005). For example, if the dyad is reluctant to further discuss end-of-life care options, this may hinder achieving the goals of that intervention element. This part of the evaluation was adapted from the Motivational Interviewing Skill Code (Miller *et al.* 2000) and consisted of three items rating the dyad's level of disclosure, cooperation, and engagement during the intervention session on a 3-point Likert scale: for example, from 1 (guarded/closed) to 3 (open) disclosure. Overall responsiveness was computed by averaging the three ratings.

**Overall quality index of intervention delivery**—This quality index was an overall evaluation of the intervention delivery after obtaining a sense of the whole intervention session by listening to the audio-recorded session, then coding and evaluating the content and process elements using transcripts, and evaluating pacing and participants' responsiveness. This overall quality index was to summarize the quality of intervention delivery using predetermined criteria. These criteria were focused on interventionist's adherence to the intervention content elements and competence in delivering the intervention elements while using the process skills, without placing too much emphasis on frequencies of particular behaviours.

This evaluation included two items – the extent to which the interventionist accomplished the goals for the five steps and performed balanced use of positive process skills and judgment in delivering the intervention on a 5-point rating scale from 1 (poor) to 5 (excellent). The first item rated the overall adherence to intervention elements by accomplishing goals for each step. Each rating was defined; for example, a rating of 5 was given when the goals of all five steps were achieved, whereas a rating of 1 was given when the goals of at least two steps and the fifth step (summarizing) were not achieved. The second item rated the overall competence (i.e. skill and judgment) of the interventionist in delivering the intervention session: for example, sensitivity to the dyad's concerns and issues, ability to focus on the topic and use of timely probing. Each rating was defined; for example, interventionists' competence was rated 5 when they utilized higher levels of clarification skills (e.g. paraphrasing or rephrasing), made smooth and timely transitions, used timely probing questions balanced between fact- and emotion-focused probing, and missed few opportunities (<3) to explore further. Interventionists' competence was rated 1 when they seldom used probing questions or clarification skills and the overall intervention session sounded like a Q&A session and frequently missed opportunities to explore or express empathy (>10). The overall quality index was computed by averaging the two ratings.

### Decisions about ratios and ratings

We set the ratios (used in the overall adherence to intervention content elements) and ratings (used in the pacing, overall dyad responsiveness and quality index) based on the analysis of previous pilot studies (Song *et al.* 2005, in press). We reviewed the intervention sessions delivered during the pilot studies to set a percentage of completed elements to be considered as satisfactory in delivery of the overall content elements. In deciding the number of points in rating, we carefully considered the following: whether each point can be discernible (whether it allows for raters to demonstrate their perceptual discriminations among the set of stimuli; Beckstead 2009), how much burden the number of points will create for the coder/rater (and, thus, whether the number is feasible), and whether such a detailed rating is useful and meaningful in monitoring fidelity. This consideration led to our decisions to use 3-point ratings on the Likert scale items for pacing and the overall dyad's receptiveness. Duration of the SPIRIT intervention typically ranged from 45 minutes to 1.5 hours. A 5-point rating would make the differences among rating points too little and, thus, not meaningful even though they are numerically discernible. The overall dyad's receptiveness was included in the assessment because of its potential impact on content fidelity. Detailed rating of the dyad, was, however, considered unnecessary and burdensome because what is evaluated and monitored should be the interventionist's, not the dyad's, behaviours.

### Application and results of the SPIRIT Intervention Fidelity Assessment Tool

To test the SPIRIT intervention, a pilot randomized controlled trial was conducted in 2006–2008 to compare the effects of the SPIRIT intervention to standard care in a sample of 58 African-American dyads: dialysis patients and their chosen surrogate decision makers. The SPIRIT study was reviewed and approved by an appropriate institutional review board at the study site.

**Participants and procedures**—Twenty-nine dyads were randomized to the intervention group and received intervention and standard care, and 29 dyads were randomized to the control group and received only standard care. There was one interventionist responsible for intervention delivery. The interventionist completed 3·5 days of competency-based training using training manuals. The training relied primarily on role playing and skill demonstration (Carroll *et al.* 2000, Hammes & Briggs 2000, Dumas *et al.* 2001, Miller *et al.* 2005). The interventionist delivered all 29 intervention sessions, which were audiorecorded and transcribed. Surveys and semi-structured interviews were used to determine the feasibility, acceptability and preliminary effects of the SPIRIT intervention on patient and surrogate outcomes (dyad congruence on goals of care, patient decisional conflict, surrogate decision-making confidence and participants' wellbeing) at 1 week and 3 months postintervention.

**Analysis**—We purposefully selected one third ( $n = 10$ ) of the 29 intervention sessions for variation in interventionist experience, including sessions from early, middle, and late phases of the study. Two trained raters independently coded and rated the 10 sessions. We then assessed inter-rater reliability using *F*-tests to estimate intraclass correlation coefficients (ICC; absolute agreement; Streiner & Norman 2003). The ICC for the four components of the fidelity tool (the overall adherence to the intervention content elements, pacing of the intervention delivery, overall dyad's responsiveness and the overall quality index of intervention delivery) ranged from 0·80 to 0·87.

**Results of the SPIRIT Intervention Fidelity Assessment Tool**—Table 3 shows an example of three intervention sessions evaluated using the tool. We selected the three intervention sessions (No. 1–3) to illustrate the full scope of ratings and how the tool may be used in training and monitoring. For example, intervention session No. 1 was the first intervention that the interventionist conducted. The ratio for the *completed* theoretical elements was 0·76 (23/30), and this less than satisfactory ( $\geq 0\cdot 80$ ) ratio was because of four *skipped* elements (ratio = 0·13), which was also less than satisfactory ( $\leq 0\cdot 10$ ). Although three elements deviated from the prescribed questions, the overall ratio for *deviated* met the satisfactory requirement ( $\leq 0\cdot 20$ ). Goals for at least two of the first four steps and Step 5 were not achieved as a result of the failure to deliver four elements and three deviated elements, and this contributed to a rating of *poor* (1) on the item of overall quality index of intervention delivery: accomplishment of goals for each step.

The total duration of the session was 40 minutes, which was shorter than the recommended total duration (48–75 minutes). Although the interventionist spent an acceptable length of time (17·1 minutes) to assess the dyad's illness representations, she spent much less time than recommended as she moved to the next steps. This was reflected in the number of *skipped* elements; three of the four were the elements of Steps 3–5. Because this was her first intervention session, she had not achieved a sense of pacing. According to her field note, she felt that she spent too much time on Step 1 and began to feel anxious to complete the rest of the steps within the recommended time. This made the session rushed; some of the important prescribed questions were skipped and, thus, the goals for those steps were not achieved.

The interventionist's overall competence in the level of skill and judgment in delivering the intervention session No. 1 was rated as *fair* (2) because of the: (i) predominant use of repeating/restating rather than the higher level skills of clarifying (43 vs. 7), (ii) use of probing questions focused more on facts than emotions (26 vs. 7), (iii) rough transitions, (iv) missed opportunities to show empathy or further explore (13) and because (v) the overall session sounded like a Q&A session. Although the dyad sounded somewhat reserved, the overall responsiveness was rated at 2·3 of 3.



The investigator informed the interventionist of the evaluation results and discussed the problem areas and plans for targeted retraining, including role play. Evaluation of intervention session Nos 2 and 3 (sessions from middle and late phases of the study) showed that the interventionist improved over time, and the levels of adherence and competence were sustained.

## Discussion

The SPIRIT Fidelity Assessment Tool was designed to evaluate interventionist adherence and competence in delivering a psycho-educational intervention to dyads. Although only one interventionist delivered the intervention sessions and was evaluated in the trial, the process of monitoring and evaluation is applicable to multiple interventionists. This tool was developed for training and monitoring an interventionist in research contexts, not for use in training clinicians. Although inter-rater reliability was acceptable (Landis & Koch 1977), the small number of sessions ( $n = 10$ ) included in the reliability assessment is a limitation. Thus, further evaluation with a larger sample of sessions will be required. However, unlike reliability evaluation in a typical psychometric evaluation framework, a  $n$  of 10 can yield large amounts of data in treatment fidelity assessment.

This form of assessment will be particularly important in examining the efficacy of the intervention on outcomes, identifying core therapeutic components of the intervention required for effectiveness, finding ways to improve the intervention, and learning how to overcome barriers and limitations to implementation of the intervention. Such a detailed assessment of an intervention may contribute to understanding of its actual (as opposed to planned) course. A psycho-educational intervention, such as the SPIRIT intervention, cannot be implemented in exactly the same way with all participants because of variations in response to individual participant needs (Perepletchikova & Kazdin 2005). For this reason, if there is no clarification between what is standardized and customized to the individual participant's needs, it will be more challenging to discern whether such variations result from interventionist drift. Our assessment tool may help us determine how much deviation or variation from the established protocol will still yield positive treatment outcomes and the intervention elements that are critical, adaptable and optional (Harshbarger *et al.* 2006).

The tool can also enhance the efficiency of training interventionists and systematic monitoring of their adherence and competence throughout the clinical trial (Orwin 2000); it can be introduced at an early stage of training and used for continuous supervision. The tool assesses both adherence to the elements of the intervention and competence in implementing the intervention. Evaluating competence of interventionists in clinical trials is necessary because an interventionist may deliver all of the content of the intervention, but in a non-therapeutic way that can contribute to low efficacy. Any deviated presentation of an intervention element and use of negative communication behaviours can be corrected and discouraged. To evaluate an hour-long single session using this tool, it typically took 2.5 hours for coding and 30 minutes for completing the tool for an experienced coder/rater. Tappin *et al.* (2000) reported that the Motivational Interviewing Skill Code could take up to 4 hours to evaluate a single session.

Questions remain concerning whether variations in intervention delivery should be assessed as part of treatment fidelity or by taking patients' characteristics into account as moderators of the intervention effects on outcomes, or both. We treated each element of the intervention as equal in importance. In the future, weights for different elements may be considered when critical and core elements have been identified.

## Conclusion

Although the intervention content elements are unique to this representational intervention, the process skill evaluation can be used for other psycho-educational interventions encompassing a general set of skills utilized in the 'talking interventions' commonly used in nursing. We have shown how researchers may go about operationalizing intervention content and process to assess interventionists' adherence and competence so that elements of the delivery of psycho-educational interventions are more visible and replicable. This tool for the systematic evaluation of intervention delivery will help to identify therapeutic components of interventions and learn how to improve them

### What is already known about this topic

- There is increasing attention to treatment fidelity in clinical trials to optimize internal validity, thereby increasing the likelihood that changes in outcomes are the result of the interventions tested.
- Systematic evaluation of treatment fidelity to cognitive-behavioural and psycho-educational interventions is challenging because of their dynamic and highly individualized nature.

### What this paper adds

- Conceptualizations of treatment fidelity tend to conflate it with study integrity.
- The specificity of individual interventions may require developing a fidelity assessment tool customized to the intervention.
- Tools for evaluating treatment fidelity of interventions with both standardized and customized components must themselves be standardized and customized.

### Implications for practice and/or policy

- To make the intervention delivery more visible and replicable, a psycho-educational intervention must be systematically evaluated using a fidelity assessment tool.
- Future research should include the examination of the relationships between the level of treatment fidelity and the targeted outcomes.

## Acknowledgments

**Funding** This work was supported by National Institutes of Health, National Institute of Nursing Research Grant No. 1R21NR009662 (M. Song) and K24-NR010244 (M.B. Happ).

## References

- Beckstead JW. Content validity is naught. *International Journal of Nursing Studies* 2009;46(9):1274–1283. [PubMed: 19486976]
- Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Ogedegbe G, Orwig D, Ernst D, Czajkowski S. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 2004;23(5):443–451. [PubMed: 15367063]
- Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, Fenton L, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence* 2000;57(3):225–238. [PubMed: 10661673]

- Chan EK, O'Neill I, McKenzie M, Love A, Kissane DW. What works for therapists conducting family meetings: treatment integrity in family-focused grief therapy during palliative care and bereavement. *Journal of Pain and Symptom Management* 2004;27(6):502–512. [PubMed: 15165648]
- Donovan HS, Ward S. A representational approach to patient education. *Journal of Nursing Scholarship* 2001;33(3):211–216. [PubMed: 11552546]
- Donovan HS, Ward SE, Song MK, Heidrich SM, Gunnarsdottir S, Phillips CM. An update on the representational approach to patient education. *Journal of Nursing Scholarship* 2007;39(3):259–265. [PubMed: 17760800]
- Dumas JE, Lynch AM, Laughlin JE, Phillips SE, Prinz RJ. Promoting intervention fidelity. Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine* 2001;20(1 Suppl.):38–47. [PubMed: 11146259]
- Forgatch MS, Degarmo DS, Beldavs ZG. An efficacious theory-based intervention for stepfamilies. *Behavior Therapy* 2005a;36(4):357–365. [PubMed: 16718303]
- Forgatch MS, Patterson GR, Degarmo DS. Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy* 2005b; 36(1):3–13. [PubMed: 16718302]
- Hammes, BJ.; Briggs, L. *Respecting Choices: Advance Care Planning Facilitator Manual*. Gundersen Lutheran Medical Foundation; La Crosse, WI: 2000.
- Harshbarger C, Simmons G, Coelho H, Sloop K, Collins C. An empirical assessment of implementation, adaptation, and tailoring: the evaluation of CDC's National Diffusion of VOICES/ VOCES. *AIDS Education and Prevention* 2006;18(4 Suppl. A):184–197. [PubMed: 16987099]
- Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *British Medical Journal* 2004;328(7455):1561–1563. [PubMed: 15217878]
- Hawe P, Shiell A, Riley T. Important considerations for standardizing complex interventions. *Journal of Advanced Nursing* 2008;62(2):267. In response to Spillane V., Byrne M.C., Byrne M., Leathem C.S., O'Malley M. & Cupples M.E. (2007) Monitoring treatment fidelity in a randomized trial of a complex intervention. *Journal of Advanced Nursing* 60(3), 343–352. [PubMed: 18394039]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33 (1):159–174. [PubMed: 843571]
- Lauver DR, Ward SE, Heidrich SM, Keller ML, Bowers BJ, Brennan PF, Kirchoff KT, Wells TJ. Patient-centered interventions. *Research in Nursing & Health* 2002;25(4):246–255. [PubMed: 12124719]
- Leeuw M, Goossens ME, de Vet HC, Vlaeyen JW. The fidelity of treatment delivery can be assessed in treatment outcome studies: a successful illustration from behavioral medicine. *Journal of Clinical Epidemiology* 2008;62(1):81–90. [PubMed: 18722086]
- Leventhal H, Friedman MA. Does establishing fidelity of treatment help in understanding treatment efficacy? Comment on Bellg *et al.* (2004). *Health Psychology* 2004;23(5):452–456. [PubMed: 15367064]
- Madson MB, Campbell TC. Measures of fidelity in motivational enhancement: a systematic review. *Journal of Substance Abuse Treatment* 2006;31(1):67–73. [PubMed: 16814012]
- Miller, WR.; Rollnick, S. *Motivational Interviewing: Preparing People to Change*. The Guilford Press; New York: 2002.
- Miller, WR.; Moyers, TB.; Ernst, D.; Amrhein, P. *Manual for the Motivational Interviewing Skill Code*. Vol. Vol. 2007. University of New Mexico; Albuquerque, NM: 2000.
- Miller WR, Moyers TB, Arciniega L, Ernst D, Forcehimes A. Training, supervision and quality monitoring of the COMBINE Study behavioral interventions. *Journal of Studies on Alcohol and Drugs* 2005;66(Suppl. 15):188–195.
- Moncher F, Prinz R. Treatment fidelity in outcome studies. *Clinical Psychology Review* 1991;11:247–266.
- Orwin RG. Assessing program fidelity in substance abuse health services research. *Addiction* 2000;95 (Suppl. 3):S309–S327. [PubMed: 11132360]
- Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: issues and research recommendations. *Clinical Psychology: Science and Practice* 2005;12:365–383.
- Resnick B, Bellg AJ, Borrelli B, Defrancesco C, Breger R, Hecht J, Sharp DL, Levesque C, Orwig D, Ernst D, Ogedegbe G, Czajkowski S. Examples of implementation and evaluation of treatment

- fidelity in the BCC studies: where we are and where we need to go. *Annals of Behavioral Medicine* 2005a;29(Suppl.):46–54. [PubMed: 15921489]
- Resnick B, Inguito P, Orwig D, Yahiro JY, Hawkes W, Werner M, Zimmerman S, Magaziner J. Treatment fidelity in behavior change research: a case example. *Nursing Research* 2005b;54(2):139–143. [PubMed: 15778656]
- Roter, DL. *The Roter Method of Interaction Process Analysis*. The Johns Hopkins University; Baltimore: 2001.
- Roter D, Larson S. The roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions. *Patient Education and Counseling* 2002;46(4):243–251. [PubMed: 11932123]
- Roter DL, Larson S, Fischer GS, Arnold RM, Tulskey JA. Experts practice what they preach: a descriptive study of best and normative practices in end-of-life discussions. *Archives of Internal Medicine* 2000;160(22):3477–3485. [PubMed: 11112242]
- Santacroce SJ, Maccarelli LM, Grey M. Intervention fidelity. *Nursing Research* 2004;53(1):63–66. [PubMed: 14726779]
- Song MK, Kirchoff KT, Douglas J, Ward S, Hammes BJ. A randomized, controlled trial to improve advance care planning among patients undergoing cardiac surgery. *Medical Care* 2005;43(10):1049–1053. [PubMed: 16166875]
- Song MK, Ward SE, Happ MB, Piraino B, Donovan HS, Shields AM, Connolly MC. Randomized controlled trial of SPIRIT: an effective approach to preparing African American dialysis patients and families for end-of-life. *Research in Nursing & Health* 2009;32:260–273. [PubMed: 19205027]
- Song MK, Donovan HD, Piraino B, Choi J, Bernardini J, Verosky D, Ward SE. Effects of an intervention to improve communication about end-of-life care among African Americans with chronic kidney disease: a pilot study. *Applied Nursing Research*. in press.
- Stein KF, Sargent JT, Rafaels N. Intervention research: establishing fidelity of the independent variable in nursing clinical trials. *Nursing Research* 2007;56(1):54–62. [PubMed: 17179874]
- Streiner, DL.; Norman, GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press; New York: 2003.
- Tappin DM, McKay C, McIntyre D, Gilour WH. A practical instrument to document the process of motivational interviewing. *Behavioral and Cognitive Psychotherapy* 2000;28:17–32.
- Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of Consulting and Clinical Psychology* 1993;61(4):620–630. [PubMed: 8370857]

**Table 1**

Example of the assessment of the intervention content elements

		<u>Interventionist adherence</u>	
		<u>Attempted</u>	<u>Completed</u>
		<u>Deviated</u>	<u>Skipped</u>
		<u>Note</u>	
Step 1. Assessing illness representations			
1	Introduction - interventionist describes the purpose of the intervention session and what kinds of questions will be asked and discussed		
2	Identity - interventionist assesses the patient's illness representations and the surrogate's representations of the patient's illness by using the following questions: "Tell me about your (your loved one's) kidney disease", "What does your (your loved one's) kidney disease bring on from day to day?" "How does your (your loved one's) illness affect your (your loved one's) daily living?"		

**Table 2**

## Example of the assessment of the process skill

Code	Total Frequency	Note (page, line number, etc.)
Repeat/restate		
Definition: skill to check for assurance or confirmation.		
Add nothing at all to what the patient or surrogate has said.		
Simply repeat it using some or all of the same words		
Paraphrase		
Definition: skill to check for assurance or confirmation.		
Change or add to what the patient or surrogate has said in a significant way to infer his or her meaning, something that he or she has not yet said directly		

**Table 3**

Evaluation of the three intervention sessions using the SPIRIT (Sharing the Patient's Illness Representations to Increase Trust) Intervention Fidelity Assessment Tool

	<u>Intervention session</u>		
	<u>No. 1</u>	<u>No. 2</u>	<u>No. 3</u>
Overall adherence to the intervention content elements*			
Elements attempted (N/A)			
Completed	0.76	0.87	0.90
Deviated	0.10	0	0
Skipped	0.13	0.13	0.10
Pacing of the intervention delivery (1–3) <sup>†</sup>			
I. Assessing illness representations	2	3	3
II. Exploring concerns	1	1	2
III. Creating condition for change	1	1	1
IV. Introducing new information	1	2	2
V. Summarizing	1	2	1
Total duration (minutes)	40	53	57
Total score (sum of ratings/5)	1.40	2.20	2.20
Overall dyad's responsiveness			
Disclosure (1, distant; 3, open)	2	3	3
Collaboration (1, resistant; 3, going along)	3	3	3
Engagement (1, passive; 3, active)	2	3	3
Total score (sum of ratings/3)	2.30	3	3
Overall quality index of intervention delivery (1–5) <sup>‡</sup>			
Accomplishment of goals for each step	1	3	4
Overall competence in the level of skill and judgment in delivering the intervention	2	2	3
Total score (sum/2)	1.50	2.50	3.50

\* Ratios (0–1).

<sup>†</sup> 1, too short; 2, acceptable; 3, too long.

<sup>‡</sup> 1, poor; 2, faire; 3, good; 4, very good; 5, excellent.