



Published in final edited form as:

J Health Care Poor Underserved. 2014 August ; 25(3): 1397–1417. doi:10.1353/hpu.2014.0132.

The Development of the DSM-5 Cultural Formulation Interview-Fidelity Instrument (CFI-FI): A Pilot Study

Neil Krishan Aggarwal, M.D., M.A.,

Research psychiatrist at the New York State Psychiatric Institute and a NIMH T32 Postdoctoral Fellow in the Department of Psychiatry at Columbia University College of Physicians and Surgeons.

Andrew Glass, M.A., M.S.,

Biostatistician at the Research Foundation for Mental Hygiene at the New York State Psychiatric Institute.

Amilcar Tirado, M.D.,

Psychiatry resident at Lincoln Medical and Mental Health Center.

Marit Boiler, M.P.H.,

Biostatistician at the New York State Psychiatric Institute.

Andel Nicasio, M.S.Ed.,

Doctoral student in clinical psychology at the University of Central Florida.

Margarita Alegría, Ph.D.,

Professor of psychology in the Department of Psychiatry at Harvard Medical School and Director of the Center for Multicultural Mental Health Research.

Melanie Wall, Ph.D., and

Professor of biostatistics in the Department of Psychiatry at Columbia University College of Physicians and Surgeons and Director of the Division of Biostatistics at the Research Foundation for Mental Hygiene at the New York State Psychiatric Institute.

Roberto Lewis-Fernández, M.D.

Professor of psychiatry in the Department of Psychiatry at Columbia University College of Physicians and Surgeons and the Director of the Center of Excellence for Cultural Competence at the New York State Psychiatric Institute.

Abstract

This paper reports on the development of the Cultural Formulation Interview-Fidelity Instrument (CFI-FI) which assesses clinician fidelity to the DSM-5 Cultural Formulation Interview (CFI). The CFI consists of a manualized set of standard questions that can precede every psychiatric evaluation. It is based on the DSM-IV Outline for Cultural Formulation, the cross-cultural assessment with the most evidence in psychiatric training. Using the New York sample of the DSM-5 CFI field trial, two independent raters created and finalized items for the CFI-FI based on

six audio-taped and transcribed interviews. The raters then used the final CFI-FI to rate the remaining 23 interviews. Inter-rater reliability ranged from .73 to 1 for adherence items and .52 to 1 for competence items. The development of the CFI-FI can help researchers and administrators determine whether the CFI has been implemented with fidelity, permitting future intervention research.

Keywords

Cultural psychiatry; cultural formulation; fidelity; cultural competence; health disparities

Health disparities for underserved racial and ethnic minorities in the United States persist along the entire pathway of mental health care. Compared with Whites, minorities are 20–50% less likely to initiate treatment^{1–5} and more likely to have shorter psychiatric visits.⁶ Minorities are also 80% less likely to receive guideline-based care⁷, 40–60% less likely to fill prescriptions^{8,9}, and 40–80% more likely to end treatment prematurely for axis I disorders than Whites.^{10–15} Disparities remain when systemic factors such as lack of medical insurance or language-matched services are accounted for in trials with subsidized, language-matched care^{16–19}, suggesting problems in patient-clinician communication. Clinicians are frequently unaware of patient cultural views around preferred illness labels²⁰, perceived illness causes^{21,22}, hesitancy to take medications^{23–26}, side effects^{27,28}, and preferences for treatments.^{29,30} Minorities who are unconvinced that clinicians understand their cultural views have seven times higher odds than Whites of ending treatment.³¹ Interventions that introduce clinicians to patient cultural views lead to increases in patient participation throughout the interview, clinician-patient information exchange, interpersonal rapport, and overall patient satisfaction.^{32–35}

To explore the patient cultural views mentioned above, cultural mental health experts created the Outline for Cultural Formulation (OCF) which appears in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders*.³⁶ The OCF assesses culture across four domains: (1) cultural identity of the individual, (2) cultural explanations of illness, (3) cultural levels of psychosocial support and functioning, and (4) cultural elements of the patient-physician relationship; a fifth domain exists for information that influences diagnosis and treatment.^{37–38} It is the most widely used cultural competence initiative in psychiatric training³⁹ and has been incorporated routinely in international outpatient clinics.^{40–42} Whereas some have used the OCF to organize material from the diagnostic assessment^{43–45}, others have used it as the basis for additional questions and identified problems with its clinical implementation. For example, the outline format of the OCF may not help clinicians formulate actual questions⁴⁶ or researchers reliably reproduce findings without a standardized intervention.⁴⁷ Guidelines on OCF use vary on its recommended length and degree of detail.^{48,49} The lack of implementation instructions has raised questions:^{50–52} Is it a separate assessment from the diagnostic interview? Should it be integrated in the standard interview through specific questions? In what service settings (inpatient or outpatient) and with what types of patients should it be conducted?

In response, the DSM-5 Cultural Issues Subgroup has revised the OCF into the Cultural Formulation Interview (CFI) for inclusion in the revised manual that was published in May 2013. In 2010 and 2011, this subgroup of cultural psychiatrists and medical anthropologists conducted literature reviews to identify the OCF's shortcomings, with revisions discussed in biweekly conference calls from March to November 2011.⁵³ The subgroup then created a standard, manualized CFI with 14 questions and probes, explanations for each question, and instructions for clinicians to use the CFI at the beginning of any diagnostic evaluation. The CFI was tested in an international field trial from November 2011 to October 2012 with the aims of examining the CFI's feasibility, acceptability, and perceived clinical utility among clinicians and patients.⁵³ Eleven collaborating sites in the United States (four sites), Canada (two), India (two), Peru (one), Kenya (one) and the Netherlands (one) coordinated all study procedures. Standardized study forms were translated into the local languages of each clinic. The CFI used in the field trial was then revised based on these outcomes into the final version included in DSM-5. For example, open-ended, semi-structured interviews with patients and clinicians after each CFI session improved the wording of certain questions and increased the total to 16 questions for the final version.⁵⁴

In making such revisions, the DSM-5 Cultural Issues Subgroup has converted the CFI into an interview through which clinicians can solicit patient cultural views in any diagnostic encounter as a clinical intervention. Interventions have been defined as purposively implemented change strategies that target malleable risk factors to produce positive outcomes.⁵⁵ The Subgroup has intended for the CFI to be used as an intervention to target the malleable risk factors of misdiagnosis, treatment disengagement, and patient dissatisfaction through change strategies of questions on patient cultural views for clinicians to incorporate in care.⁵⁶ Researchers have criticized cultural competence intervention developers for not assessing clinician adherence and competence after being trained in such interventions through quantitative, reproducible instruments.⁵⁷ The lack of adherence and competence measures for clinician cultural competence interventions is a substantial gap in the literature that this project seeks to address. Assessing clinician adherence and competence as indices of intervention fidelity are important throughout intervention research in order to revise training, monitor delivery, and revise the intervention based on clinician and patient responsiveness.^{58,59} By intervention fidelity, we specifically mean "the extent to which core components of interventions are delivered."⁵⁹ Adherence measures can help determine that an intervention is distinguishable from others.^{60,61} Competence measures can determine the skill with which an intervention is implemented.^{62,63} If the global circulation of DSM-IV is an indication^{64,65}, then DSM-5 may also circulate widely, making an instrument that measures clinician fidelity to the CFI of timely and topical interest. This paper reports initial CFI clinician adherence and competence data through the Cultural Formulation Interview-Fidelity Instrument (CFI-FI) based on the New York sample of the DSM-5 CFI field trial. Our aims were to develop the CFI-FI, explore possible differences in CFI fidelity based on clinician demographics, and investigate reasons for discrepant findings among raters.

Methods

The DSM-5 CFI field trial

The DSM-5 Cultural Issues Subgroup, led by the New York State Psychiatric Institute (NYSPI) and Columbia University, designed the field trial. NYSPI and the outpatient clinics of New York Presbyterian Hospital and the Washington Heights Community Service formed the New York site. The aims of the overall field trial were to assess the feasibility, acceptability, and perceived clinical utility of the CFI among patients and clinicians. These aims were assessed through quantitative scales and semi-structured qualitative interviews in order to refine the CFI for DSM-5. Each site agreed to enroll a minimum of 30 psychiatric outpatients. All patients were referred by regular clinic staff not part of the study. New patients were referred with the patient's suspected DSM-IV diagnosis and existing patients were referred with the patient's DSM-IV diagnosis at the time of the study. The patient was interviewed by a study clinician and the interview was structured as an intake assessment. The interview consisted of a one-hour, diagnostic interview that began with the clinician conducting the CFI after which the study clinician transitioned into a diagnostic evaluation. After the CFI-enhanced diagnostic interview, patients and clinicians completed quantitative instruments (developed for this study) separately about their experiences with the CFI. Clinicians also completed a form that mirrored the referral form, documenting the patient's diagnoses and recommended treatment plan. Clinicians and patients then participated in separate fifteen-minute, semi-structured interviews with research staff to discuss their experiences with the CFI-enhanced evaluation. All CFI sessions and post-session interviews were audio-taped. The Institutional Review/Ethics Board at each site approved the study with patients and clinicians providing written consent. Patients were paid \$35 for participation, but clinicians were not paid.

Patients

Eligible patients were of any race or ethnicity; between 18 and 80 years old; fluent in a language spoken by the clinicians at each local site; and with any psychiatric diagnosis. Patients were excluded if found by study or referring clinicians during the clinical examination to be acutely suicidal or homicidal; intoxicated or in substance withdrawal; or with any interview-interfering condition such as dementia, mental retardation, or florid psychosis.

Clinicians

Eligible clinicians were on staff at each site and possessed a terminal clinical degree (MD, MSW, PhD) permitting them to practice independently. Clinicians were recruited by research staff at each site through fliers, word-of-mouth, and in-person presentations. Interested clinicians were asked to attend a 90-minute training session consisting of reviewing CFI guidelines, watching a video of a mock CFI session, and role-playing simulations. The training protocol employed these three modalities as suggested by intervention researchers to increase clinician fidelity.⁵⁸ Each clinician was asked to conduct between three and six interviews. A minimum of three interviews was established to insure a range of experiences in using the CFI so that clinicians would not be "CFI-naïve" for every case. At the same time, clinicians were limited to six interviews in order to maximize the

participation of interested study clinicians. Clinicians were excluded if they could not attend the training. Study clinicians did not interview their own patients since their extant knowledge about the patient could confound the aim of testing the CFI's clinical utility.

The Cultural Formulation Interview (CFI)

The field trial CFI [available upon request] included 14 questions with instructions to ask the questions in order based on the four domains of the DSM-IV OCF. The manualized CFI included questions to patients in the right column and instructions to clinicians about the purpose of each question in the left column. The CFI topics were: (1) the patient's presenting problem, (2) the problem's perceived level of severity, (3) a description of the problem, (4) the patient's preferred idiom of distress, (5) perceived illness causes, (6) the role of the patient's social network in worsening and improving the problem, (7) the relationship of cultural identity to improving or worsening the problem, (8) methods of self-coping, (9) past help seeking on the most and least useful treatments, (10) cultural barriers to care (e.g., Stigma and discrimination), (11) concerns about the clinician's background, and (12) current treatment preferences. The exact wording of the questions assessing these topics can be found elsewhere.⁵⁴ Topics (4) and (9) included probe questions for additional information from patients, and these two probes were not included in the CFI-FI since they did not represent new topics.

The New York sample

Table 1 includes the characteristics of the New York site. Patient ages ranged from 29 years to 79 years and clinician ages ranged from 29 to 54 years. Thirty-two patients were recruited from the New York Presbyterian Hospital and the New York State Psychiatric Institute from November 2011 to May 2012. Patients were mostly female (n=22) and Hispanic/Latino (n=22). No patients met criteria for exclusion by study clinicians. Eleven clinicians attended the training and seven enrolled in the study. The seven study clinicians were mostly women (n=5), Hispanic/Latino (n=3), and psychiatrists (n=3). All clinicians completed three or more interviews.

Development of the CFI-FI

Given the variety of languages spoken in the field trial across all sites and restrictions on data sharing of patient audio recordings, items for the CFI-FI were generated from only the New York sample. Fidelity items should be generated with as much source material as possible so that instrument developers can work with the most amount of available information.^{66,67} For studies with sample sizes of 30 and below, fidelity instruments should be developed through session transcriptions that provide maximal source material when used with audio recordings for ratings.⁶⁷ Therefore, items were generated by listening to taped CFI sessions, analyzing professionally-transcribed CFI sessions, and examining the manualized CFI document provided to all clinicians to maximize rating accuracy.

Systematic reviews of the intervention development literature have provided definitions for five components of intervention fidelity: (1) clinician adherence to methods, (2) clinician competence, (3) participant responsiveness, (4) intervention distinctness, and (5) dosage of

the intervention⁶⁸; these were adapted to create CFI-FI items which appear in the tables below.

Adherence to methods—Adherence to methods has been defined as the extent to which the implementation of intervention activities and methods corresponds with written instructions.⁶⁸ Adherence measures requiring raters to make “present-or-absent” distinctions are economical and more likely to produce reliable ratings than Likert scales.⁶⁷ Adherence was operationalized by assessing whether the clinician asked CFI questions for each of the 12 topics (zero=no/absent and one=yes/present). All questions in this subscale begin with the question stem: “Did the clinician ask a CFI question about” [the CFI topic] to focus on clinician behavior.

Competence—Competence (also called “quality of delivery”) has been defined as the extent to which a provider delivers intervention content in a way that is considered ideal by intervention developers.⁶⁸ An intervention can be conceived of as a collection of activities and methods that are: (1) unique and essential to the intervention, (2) essential but not unique to the intervention (common to other interventions), (3) acceptable but not necessary for the intervention, and (4) proscribed, with competence focusing on areas (1) and (2).⁶⁷ CFI elements that were emphasized as *unique and essential* during training sessions include: patient-centeredness, clinician word matching of patient terms, and the clinician’s ability to elicit illness narration from the patient. While these elements may be ideal characteristics of other types of interviews and interventions, they took special form within the CFI. For example, patient centeredness was operationalized as the clinician not arguing, confronting, or correcting the patient since the CFI introduction states that there are no right or wrong answers.⁵⁶ Clinician word matching comes from the anthropological literature on idioms of distress which are the cultural terms that people use to convey sickness.⁶⁹ Clinician word matching has been shown to increase patient satisfaction⁷⁰, and the clinician is instructed to use the patient’s terms for sickness throughout the CFI. For example, clinicians should use “feeling down” as an idiom if mention by the patient rather than “experiencing major depression.” Illness narration refers to the clinician’s skill at eliciting the patient’s experience of illness through open-ended questions that invite narrative rather than asking close-ended questions about diagnostic symptoms.³⁸ CFI ideals that are *essential but not unique* include empathy and clarification. Empathy has been difficult to define and rate reliably, so we used the concept of the clinician making statements that paraphrased or named the patient’s emotional state.⁷¹ Empathy is distinct from word matching given that paraphrases are permitted to rate empathy. Clarification refers to the clinician’s use of probe questions to understand patient responses that may have appeared unclear. Clinicians were encouraged during training to pay attention to these competencies. The competencies were also emphasized in the CFI guidelines provided to all clinicians during the session. For all items, raters considered if competencies were delivered: never (zero), less than half (<50%) of the CFI session (one), and at least half (≥ 50%) of the CFI session (two) as a whole. We selected a three-point Likert scale for rating since Likert scores with more than three anchor points may reduce inter-rater reliability.⁷² Raters measured competence for the whole CFI session since competence is best measured when the entire session is taken into account.⁶⁷

Participant responsiveness—Participant responsiveness has been defined as the extent to which participants are engaged by and involved in the intervention.⁶⁸ Participant responsiveness can affect clinician adherence and competence to an intervention when clinicians perceive that participants are not responding as intended; clinicians may then modify intervention delivery accordingly.⁶⁸ In coding interviews with the first draft of the CFI-FI, we observed that patients with thought disorders did not answer in intended ways, resulting in a variety of clinician responses such as interruptions, omitted questions, or accelerated interviews. Therefore, we operationalized participant responsiveness by assessing whether: a) the patient responded to the CFI question (“Did the patient respond to the question?”), and b) the response was relevant to the topic asked (“Was the patient’s response about [the CFI topic]?”). We termed a relevant response as a “good response.” Each item was coded on a no (zero) or yes (one) basis to correspond with clinician adherence.

Intervention distinctness—Intervention distinctness has been defined as those components which must be excluded to insure the uniqueness of an intervention so that it can be reliably differentiated from other interventions.⁶⁸ Intervention distinctness focuses on making sure that undesired and proscribed activities and methods are not delivered lest they endanger the intervention.⁶⁷ For the CFI-FI, intervention distinctness was focused on two rating items: drift and order. Drift was operationalized as the extent to which clinicians interspersed topics from the standard diagnostic interview (symptom assessments from the history of present illness, current medications, detailed medical history, family history, and mental status examination) vs. focusing on CFI topics. Order was the extent to which clinicians asked CFI questions serially as instructed in the written guidelines. To distinguish this construct from adherence to method, order was rated by assessing whether all of the topics asked were asked serially. In other words, order was rated even if all CFI topics were not asked. Both items were developed based on calls in the literature that the cultural formulation should serve as a separate, stand-alone interview from the standard diagnostic interview.^{49–51} Items were rated on the three-point Likert scale of zero, one, and two as described above.

Dosage—Dosage has been defined as the amount of intervention received by participants.⁶⁷ We operationalized dosage by quantifying the total length of time of the CFI session through (minute:second) format. This measurement was intended to provide additional information beyond the adherence measure of number of questions that clinicians asked. Theoretically, clinicians could ask all questions very quickly or slowly, and we wanted to examine the length of CFI sessions.

Rater guidelines

A detailed rater’s manual [available upon request] for the CFI was developed to maximize a consistent and reliable approach to ratings. The manual introduces the purpose of the rating instrument, encourages raters to read the instrument in its entirety for content familiarity, and provides instructions for rating each item based on observable behaviors. The manual also details the purpose of each item, providing examples and counter-examples to avoid ambiguous ratings.

Rater selection and procedure

In developing instruments to rate clinician fidelity to a new intervention, a pool of raters ideally includes clinicians who are trained and untrained in the intervention to prevent potential bias in inter-rater reliability based on prior exposure.⁷³ Two raters, physicians bilingual in English and Spanish, rated all data independently by listening to audiotapes and reading transcripts simultaneously. One rater was fully trained in the CFI as an advisor to the DSM-5 Cultural Issues Subgroup and helped to oversee the CFI field trial across all sites (NKA). The other rater was fully untrained in the CFI and worked as a volunteer (AT). Of 32 sessions, fifteen were in English and seventeen were in Spanish. Technical difficulties prevented three sessions from recording, leaving 29 sessions that were audio-taped and transcribed. Consistent with general methods in intervention research^{58, 61}, the first author (NKA) randomly selected 20% of the New York sample (six interviews) to create and finalize items for clinician adherence and competence. These items were used to rate the rest of the sample (23 interviews).

During development of the CFI-FI from October to December 2012, both raters met three times to discuss ratings made independently before the meeting and their use of the CFI-FI. These meetings also served to act as rater “calibration” sessions to assess reliability, develop consensus ratings, discuss reasons for variant ratings, and clarify CFI-FI instructions and items. Because item instructions, questions, and anchor points underwent changes, these ratings are not included in this paper. Based on this pilot work, the CFI-FI was finalized and then formally tested. In January and February 2013, the raters used the final CFI-FI version to rate the remaining 23 CFI interviews independently by listening to audiotapes and reading transcripts simultaneously. Both raters rated all 23 tapes included in the present analyses. The raters did not meet at all while rating the remaining 23 CFI interviews to avoid biasing subsequent ratings. After both raters completed rating all 23 interviews, they met during the last week of February 2013 to discuss their findings. All ratings were made before this week and were not changed subsequently.

For each meeting during the pilot and final rating phases, raters entered their scores independently on a Microsoft Excel spreadsheet that was coded to the CFI-FI. Two columns were created to list the ratings of each rater next to the row for each CFI-FI item. A third column titled “Consensus” was created during the meeting by both raters to discuss all item scores. In situations where the raters agreed, the consensus score was the same as both scores given by the raters. In situations where the raters disagreed, raters discussed reasons for why their scores were discrepant and finalized one consensus score for that item. The final consensus score was not used at all during quantitative analyses, but acted as a data source for qualitative analyses on rater disagreement after creation of the final CFI-FI. In addition, an audit trail of analytical memos and meeting notes was used to insure rigor and validity of data analyses.⁷⁴ The CFI-FI, spreadsheets, memos, and notes in all phases are available upon request.

Statistical Analyses

For items linked to each CFI topic such as scores for clinician adherence to the question and patient responsiveness (any response, good response), sub-scales for each CFI question

(asked, response, good response) were computed by summing the number of possible positive responses (range zero to twelve). For example, a score of twelve on the asked subscale means that every CFI topic was asked by the clinician during the interview. Percent of positive responses was calculated as the total of “yes” (one) responses to the CFI questions divided by twelve, the number of CFI topics. As a separate statistic, percent concordance was defined as the number of items that were rated the same by both raters divided by the total number of asked items. To assess inter-rater reliability, percent concordance was used for individual items, and correlations were used for sub-scales. Percent concordance was preferred over the *kappa* statistical test because *kappa* is not recommended when the prevalence of outcomes is too low (<.05) or too high (>.95)⁷⁵, as for most items in our study. Overall or average percent concordance for raters is not presented since rater concordance is linked to constructs that are rated differently for patients and clinicians. Frequencies for all adherence, patient responsiveness, competence, and intervention distinctiveness scores are presented based on the consensus scores of both raters.

To assess correlations among clinician adherence, clinician competence, patient responsiveness, intervention distinctness, and dosage, Spearman’s rank correlation coefficient was used. Spearman’s correlation measures the relationship between two variables with measures taken for ordinal scales (for example, measurements for competencies such as “empathy”) by converting each variable to ranks (as in our Likert scales) to measure whether the ranks of one variable co-vary with the ranks of another variable.⁷⁶

Results

Adherence to method

Adherence was measured through the *asked* sub-scale in which raters scored each item as zero (not asked) or one (asked).

Overall, clinicians asked most CFI questions, with the percentage of positive responses under the asked subscale ranging from 87% to 100% (Table 2). Percent concordance between raters ranged from 73.9% to 100% (Table 2), indicating high inter-rater reliability. Table 3 presents the frequencies of sum scores for adherence to method divided by the *asked, responded, and good response* subscales. The highest possible fidelity score for each subscale was twelve, corresponding to total number of CFI topics. Overall frequency refers to the number of scores for each subscale, totaling to a maximum of 23 (n=23). Frequencies are also shown by language and clinician type. Sixteen of 23 subjects (69.6%) were asked all twelve questions, as indicated by the “fidelity score.” Next, we assessed whether the adherence subscales differed by language or type of clinician conducting the interview. Overall, adherence did not differ substantially by language of original interview or type of study clinician. The correlation of sum scores between raters was .43, indicating that the clinician adherence subscale of the CFI-FI was largely independent of raters (results not shown, but available from the authors).

Patient responsiveness

Patient responsiveness was measured through the *response* and *good response* sub-scales (Table 2). For the response subscale (not shown), raters scored each item as zero (the patient did not respond) or one (the patient responded). For the good response subscale, raters similarly scored each item as zero (the response was not relevant) or one (the response was relevant). Overall, patients responded and provided good responses to most CFI questions, with each response or good response item (scored a one) occurring at least 87% of the time (Table 2). In all instances where clinicians asked the CFI question, patients responded. Concordance between raters ranged from 73.9% to 100% (Table 2), indicating high inter-rater reliability. A total of 16 out of 23 subjects (69.6%) had any response to all 12 questions and 15 of 23 subjects (65.2%) had a good response to all 12 questions (Table 3).

The correlation of sum score between raters was .43 for any response and .63 for good response, indicating that the any response subscale was mildly correlated between raters and the good response subscale was moderately correlated. Overall, any patient response did not differ substantially by language or clinician type (Table 3). A good response was more frequently associated with non-MD clinicians than with MD clinicians (10 patients vs. 5 patients) in those instances when patients provided a good response for all CFI questions (score=12). Non-MD clinicians were able to elicit a good response for all CFI questions (score =12) than were MD clinicians – for example, 71% of patients (10 out of 14) meeting with non-MD clinicians gave all good responses vs. 56% of patients (5 out of 9) interviewed by MD clinicians. This difference did not reach statistical significance due to the small sample size.

Clinician competence

Clinician competence was measured by the items listed in Table 4 in which raters scored each item as zero (did not occur), one (occurred less than 50% of the CFI session) and two (occurred 50% or more of the CFI session).

Overall, clinicians were judged to deliver the CFI with relatively high competence, with raters scoring that clarification and illness narration competencies occurred in 50% or more of the CFI session for all sessions. Percent concordance between raters demonstrated a wide range of variability, from 52.2% for word matching to 100% for clarification, indicating moderately high to extremely high levels of inter-rater reliability.

Intervention distinctiveness

Intervention distinctiveness was measured by items on drift and order (Table 4). Drift was rated as never occurring in 21 of 23 subjects (91.3%), indicating that clinicians covered only CFI topics during the designated CFI session. Order was rated as 100%, indicating that CFI questions were asked serially in all 23 rated sessions. Concordance between raters for drift and order was 86.2% and 100%, respectively, indicating high levels of inter-rater reliability.

Dosage

CFI session length averaged 24.48 minutes (SD=9.52). English and Spanish-language interviews averaged 23.02 minutes (SD=6.47) and 25.94 minutes (SD=11.98), respectively.

Interviews by MDs and non-MDs averaged 20.44 minutes (SD=6.52) and 26.78 minutes (SD=10.37), respectively. No significant correlations were found between session length and sum scores on the *asked* ($r=.04$, $p=.86$), *responded* ($r=.04$, $p=.86$), or *good response* scales ($r=.005$, $p=.98$).

Correlation among variables

Table 5 presents the Spearman rank correlation coefficients (r) for all five constructs of intervention fidelity measured through the CFI-FI.

Statistically significant results are bolded. Adherence to method and patient responsiveness were highly correlated. Adherence to method was completely correlated with the any response subscale ($r=1$, $p<.0001$) and was strongly correlated with the good response subscale ($r=.92$, $p<.0001$). That is, patients answered every question, usually providing relevant answers. Word matching, one component of clinician competence, and dosage were moderately inversely correlated ($r=-.53$, $p=.01$), indicating that a higher frequency of clinician-patient word matching was associated with shorter CFI sessions. No other statistically significant correlations were found between other constructs. Because Spearman rank correlation tests the correlation between ordinal variables that are internally distinguished by a ranking system, variables in which observed values are all rated the same (all CFI sessions are rated exactly the same by both raters) cannot be internally ranked. Since clarification, illness narration, and order items were rated the same for all 23 subjects (Table 4), Spearman coefficients are not reported for these items.

Discussion

To our knowledge, this is the first paper to report an empirically derived and tested fidelity instrument for a clinician cultural competence measure. We found that clinicians adhered to all twelve topics assessed by the CFI at least 87% of the time. However, the raters consistently achieved lower concordance of 73.9% in rating whether clinicians asked, patients responded, and patients provided a good response to the CFI topic on the severity of the problem. Rater discrepancy emerged from the clinician's exact method of asking the question. One rater rated items positively if the clinician asked "What troubles you most about the problem?"; the other rater did not insist "most" being included in the question. For patient responsiveness, 69.6% of subjects provided a response and 65.2% of subjects had a good response to questions on all twelve topics. For intervention distinctiveness, drift only occurred in 8.7% of sessions and order was rated as occurring in all patients. It is possible that our convenience sample of clinicians was biased toward those completing the CFI as intended. However, the training protocol of guideline review, video demonstration, and role-playing exercises based on recommended models of intervention development⁵⁸ may have increased clinician fidelity with the CFI. That percent concordance between raters ranged from 52% to 100% for all measures – with most concordance greater than 60% - suggests high CFI-FI inter-rater reliability.⁷⁷

Clinician competence and dosage demonstrated unanticipated findings. Word matching and dosage were moderately inversely correlated, suggesting that sessions in which clinicians matched patient illness terms completed the CFI in shorter time. This may be due to greater

patient-clinician understanding during the session, though future studies would need to test this hypothesis. Dosage also differed by demographics. Spanish-language sessions were longer than those in English and sessions conducted by non-MDs were 7 minutes longer than MD-conducted sessions. Future work can examine if CFI length affects implementation by type of clinician: those clinicians reimbursed by time in service settings may be more likely to implement the CFI.

This study has some limitations. First, we excluded three interviews in our sample due to recording error. The exclusion of these data could have affected inter-rater reliability. Nonetheless, the sample size ultimately used for rating (n=23) is twice the suggested sample size of 10–12 patients recommended for pilot studies of fidelity instruments in intervention development.⁵⁸ Second, our method of rating audiotapes and reading transcripts simultaneously may not be generalizable outside of research settings. The cost of obtaining transcriptions often prevents intervention developers from using an entire dataset, and many must resort to analyzing snippets of interviews that are randomly selected.⁶⁷ However, we sought to develop the CFI-FI using all available data at the New York site. Future studies could test differences in ratings based on listening to audiotapes, reading transcripts, or direct observation. Third, our method of measuring inter-rater reliability by percent concordance does not control for instances of chance agreement. Still, our data analyses demonstrate robust high inter-rater reliability for all constructs which cannot be entirely due to chance. Different raters with larger sample sizes can test whether the inter-rater reliability reported here is due to consistent rating with the CFI-FI or from chance. Fourth, we have not controlled for patient diagnoses or illness severity in our dimension of participant responsiveness. Our measurements of participant responsiveness may confound patient illness severity with lack of clinician adherence to CFI training. However, the primary goal of a pilot fidelity instrument is to achieve high inter-rater reliability to determine score reproducibility.^{58,61} Concordance between raters for participant responsiveness ranged from 73.9% to 100% (Table 2), indicating high inter-rater reliability. Future studies can disentangle whether lower scores for participant responsiveness are due to patient illness severity or CFI training. Fifth, our lowest percent concordance for the clinician competence measure was for word matching (52.2%). Raters could not achieve higher inter-reliability, possibly because patients reported multiple terms and raters selected different terms to rate word matching. Finally, we developed the CFI-FI based on the CFI version used in the DSM-5 field trial. Criticisms could be raised that this CFI-FI is not generalizable to the CFI version ultimately included in DSM-5. Nonetheless, intervention development research, by definition, includes modifications to interventions, training protocols, and fidelity measures based on scientific stage of development; just as interventions change, so should fidelity measures since “one size cannot fit all.”⁷⁸ The CFI-FI can be modified to match the final CFI in DSM-5 to include items on clinician adherence and patient responsiveness, and our work suggests a question format and item rating scales for potential modifications.

Despite these potential limitations, the CFI-FI advances the science of clinician cultural competence. Systematic reviews have noted the lack of quantitative studies assessing cultural competence training⁵⁷ and whether clinician cultural competence training improves patient outcomes.⁷⁹ The CFI can be used as an intervention that builds from the evidence base of cultural competence used in psychiatric training. Another line of work can examine

the potential modifications needed to adapt the field trial CFI-FI for the final CFI version included in DSM-5. As sample sizes increase for ratings, the CFI-FI can be tested for convergent and divergent validity with other scales such as the Agency for Healthcare Research and Quality's Cultural Competence Item Set of the Consumer Assessment of Healthcare Providers and Systems.⁷² Finally, future research can examine how various training modalities affect CFI fidelity; how CFI fidelity varies by clinician, patient, and rater demographics; and whether rating methods affect inter-rater reliability. In this regard, it will be important to understand if and how clinicians, administrators, and health systems can assess their cultural competence initiatives through faithful application of the CFI, and along with it, the CFI-FI, to determine clinician fidelity.

Acknowledgements

The first author is supported by a NIMH T32 post-doctoral fellowship in clinical research. He is an official advisor within the DSM-5 Cultural Issues Subgroup of which the senior author is the chairperson. The authors acknowledge a grant from the American Psychiatric Association for the DSM-5 field trial. The first author also received a Policy Scholar Award grant from the New York State Office of Mental Health to develop the CFI-FI.

References

1. Sirey JA, Meyers BS, Bruce ML, et al. Predictors of antidepressant prescription and early use among depressed outpatients. *Am J Psychiatry*. 1999 May; 156(5):690–696. [PubMed: 10327900]
2. Sentell T, Shumway M, Snowden L. Access to mental health treatment by English language proficiency and race/ethnicity. *J Gen Intern Med*. 2007 Nov; 22(Suppl 2):289–293. [PubMed: 17957413]
3. Alegría M, Chatterji P, Wells K, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. *Psychiatr Serv*. 2008 Nov; 59(11):1264–1272. [PubMed: 18971402]
4. Lê Cook B, McGuire TG, Lock K, et al. Comparing methods of racial and ethnic disparities measurement across different settings of mental health care. *Health Serv Res*. 2010 Jun; 45(3):825–847. [PubMed: 20337739]
5. Gonzalez A, Weersing VR, Warnick EM, et al. Predictors of treatment attrition among an outpatient clinic sample of youths with clinically significant anxiety. *Adm Policy Ment Health*. 2011 Sep; 38(5):356–367. [PubMed: 20976618]
6. Olfson M, Cherry DK, Lewis-Fernández R. Racial differences in visit duration of outpatient psychiatric visits. *Arch Gen Psychiatry*. 2009 Feb; 66(2):214–221. [PubMed: 19188544]
7. Mallinger JB, Fisher SG, Brown T, et al. Racial disparities in the use of second-generation antipsychotics for the treatment of schizophrenia. *Psychiatr Serv*. 2006 Jan; 57(1):133–136. [PubMed: 16399976]
8. Chen J, Rizzo JA. Racial and ethnic disparities in antidepressant drug use. *J Ment Health Policy Econ*. 2008 Dec; 11(4):155–165. [PubMed: 19096090]
9. Busch AB, Huskamp HA, Neelon B, et al. Longitudinal racial/ethnic disparities in antimanic medication use in bipolar-I disorder. *Med Care*. 2009 Dec; 47(12):1217–1228. [PubMed: 19786909]
10. Warden D, Trivedi MH, Wisniewski SR, et al. Predictors of attrition during initial (citalopram) treatment for depression: a STAR*D report. *Am J Psychiatry*. 2007 Aug; 164(8):1189–1197. [PubMed: 17671281]
11. Busch AB, Lehman AF, Goldman H, et al. Changes over time and disparities in schizophrenia treatment quality. *Med Care*. 2009 Feb; 47(2):199–207. [PubMed: 19169121]
12. Kreyenbuhl J, Nossel IR, Dixon LB. Disengagement from mental health treatment among individuals with schizophrenia and strategies for facilitating connections to care: a review of the literature. *Schizophr Bull*. 2009 Jul; 35(4):696–703. [PubMed: 19491314]

13. Fortuna LR, Alegria M, Gao S. Retention in depression treatment among ethnic and racial minority groups in the United States. *Depress Anxiety*. 2010 May; 27(5):485–494. [PubMed: 20336808]
14. Gerber BS, Cho YI, Arozullah AM, et al. Racial differences in medication adherence: a cross-sectional study of Medicare enrollees. *Am J Geriatr Pharmacother*. 2010 Apr; 8(2):136–145. [PubMed: 20439063]
15. Kozhimannil KB, Trinacty CM, Busch AB, et al. Racial and ethnic disparities in postpartum depression care among low-income women. *Psychiatr Serv*. 2011 Jun; 62(6):619–625. [PubMed: 21632730]
16. Arnow BA, Blasey C, Manber R, et al. Dropouts versus completers among chronically depressed outpatients. *J Affect Disord*. 2007 Jan; 97(1–3):197–202. [PubMed: 16857266]
17. Sánchez-Lacay JA, Lewis-Fernández R, Goetz D, et al. Open trial of nefazodone among Hispanics with major depression: efficacy, tolerability, and adherence issues. *Depress Anxiety*. 2001; 13(3): 118–124. [PubMed: 11387731]
18. Warden D, Rush AJ, Wisniewski SR, et al. What predicts attrition in second step medication treatments for depression?: a STAR*D Report. *Int J Neuropsychopharmacol*. 2009 May; 12(4): 459–473. [PubMed: 18611293]
19. Lewis-Fernández R, Balán IC, Patel SR, et al. Impact of motivational pharmacotherapy on treatment retention among depressed Latinos. *Psychiatry*. in press.
20. Lewis-Fernández R, Das AK, Alfonso C, et al. Depression in US Hispanics: diagnostic and management considerations in family practice. *J Am Board Fam Pract*. 2005 Jul-Aug;18(4):282–296. [PubMed: 15994474]
21. Stewart SM, Simmons A, Habibpour E. Treatment of culturally diverse children and adolescents with depression. *J Child Adolesc Psychopharmacol*. 2012 Feb; 22(1):72–79. [PubMed: 22251021]
22. Schraufnagel TJ, Wagner AW, Miranda J, et al. Treating minority patients with depression and anxiety: what does the evidence tell us? *Gen Hosp Psychiatry*. 2006 Jan-Feb;28(1):27–36. [PubMed: 16377362]
23. Burnett M, Genao I, Wong WF. Race, culture, and trust: why should I take a shot if I'm not sick? *Ethn Dis*. 2005 spring;15(2 Suppl 3) S3-13-S3-6.
24. Gonzalez JM, Alegría M, Prihoda TJ, et al. How the relationship of attitudes toward mental health treatment and service use differs by age, gender, ethnicity/race and education. *Soc Psychiatry Psychiatr Epidemiol*. 2011 Jan; 46(1):45–57. [PubMed: 19921079]
25. Givens JL, Houston TK, Van Voorhees BW, et al. Ethnicity and preferences for depression treatment. *Gen Hosp Psychiatry*. 2007 May-Jun;29(3):182–191. [PubMed: 17484934]
26. Cooper LA, Gonzales JJ, Gallo JJ, et al. The acceptability of treatment for depression among African-American, Hispanic, and white primary care patients. *Med Care*. 2003 Apr; 41(4):479–489. [PubMed: 12665712]
27. Gadkari AS, McHorney CA. Medication nonfulfillment rates and reasons: narrative systematic review. *Curr Med Res Opin*. 2010 Mar; 26(3):683–705. [PubMed: 20078320]
28. Interian A, Martinez I, Rios LI, et al. Adaptation of a motivational interviewing intervention to improve antidepressant adherence among Latinos. *Cultur Divers Ethnic Minor Psychol*. 2010 Apr; 16(2):215–225. [PubMed: 20438160]
29. Villagran M, Hajek C, Zhao X, et al. Communication and culture: predictors of treatment adherence among Mexican immigrant patients. *J Health Psychol*. 2012 Apr; 17(3):443–452. [PubMed: 21900335]
30. Snipes SA, Sellers SL, Tafawa AO, et al. Is race medically relevant? A qualitative study of physicians' attitudes about the role of race in treatment decision-making. *BMC Health Serv Res*. 2011 Aug 5.11:183. [PubMed: 21819597]
31. van Wieringen JC, Harmsen JA, Bruijnzeels MA. Intercultural communication in general practice. *Eur J Public Health*. 2002 Mar; 12(1):63–68. [PubMed: 11968523]
32. Newes-Adeyi G, Helitzer DL, Roter D, et al. Improving client-provider communication: evaluation of a training program for women, infants and children (WIC) professionals in New York state. *Patient Educ Couns*. 2004 Nov; 55(2):210–217. [PubMed: 15530756]

33. Roter DL, Wexler R, Naragon P, et al. The impact of patient and physician computer mediated communication skill training on reported communication and patient satisfaction. *Patient Educ Couns.* 2012 Sep; 88(3):406–413. [PubMed: 22789149]
34. Tibaldi G, Salvador-Carulla L, García-Gutierrez JC. From treatment adherence to advanced shared decision making: new professional strategies and attitudes in mental health care. *Curr Clin Pharmacol.* 2011 May; 6(2):91–99. [PubMed: 21592062]
35. Pawlikowska T, Zhang W, Griffiths F, et al. Verbal and non-verbal behavior of doctors and patients in primary care consultations - how this relates to patient enablement. *Patient Educ Couns.* 2012 Jan; 86(1):70–76. [PubMed: 21621365]
36. Mezzich JE, Kirmayer LJ, Kleinman A, et al. The place of culture in DSM-IV. *J Nerv Ment Dis.* 1999 Aug; 187(8):457–464. [PubMed: 10463062]
37. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders.* fourth edition. Arlington: American Psychiatric Publishing, Inc; 2000.
38. Lewis-Fernández R. Cultural formulation of psychiatric diagnosis. *Cult Med Psychiatry.* 1996 Jun; 20(2):133–144. [PubMed: 8853962]
39. Aggarwal NK, Rohrbaugh RM. Teaching cultural competency through an experiential seminar on anthropology and psychiatry. *Acad Psychiatry.* 2011 Fall;35(5):331–334. [PubMed: 22007094]
40. Kirmayer LJ, Thombs BD, Jurcik T, et al. Use of an expanded version of the DSM-IV outline for cultural formulation on a cultural consultation service. *Psychiatr Serv.* 2008 Jun; 59(6):683–686. [PubMed: 18511590]
41. Bäärnhielm S, Scarpinati Rosso M. The cultural formulation: a model to combine nosology and patients' life context in psychiatric diagnostic practice. *Transcult Psychiatry.* 2009 Sep; 46(3):406–428. [PubMed: 19837779]
42. Rohlf H, Knipscheer JW, Kleber RJ. Use of the cultural formulation with refugees. *Transcult Psychiatry.* 2009 Sep; 46(3):487–505. [PubMed: 19837783]
43. Lu FG, Lim RF, Mezzich JE. Issues in the assessment and diagnosis of culturally diverse individuals. *American Psychiatric Press Review of Psychiatry.* 1995; 14:477–510.
44. Lim RF, Diamond RJ, Chang JB, et al. Using non-feature films to teach diversity, cultural competence, and the DSM-IV-TR outline for cultural formulation. *Acad Psychiatry.* 2008 Jul-Aug; 32(4):291–298. [PubMed: 18695030]
45. Lim, RF. *Clinical manual of cultural psychiatry.* Arlington: American Psychiatric Publishing, Inc; 2006.
46. Lewis-Fernández R. The cultural formulation. *Transcult Psychiatry.* 2009 Sep; 46(3):379–382. [PubMed: 19837777]
47. Alarcón RD. Culture, cultural factors and psychiatric diagnosis: review and projections. *World Psychiatry.* 2009 Oct; 8(3):131–139. [PubMed: 19812742]
48. Group for the Advancement of Psychiatry. *Cultural assessment in clinical psychiatry.* Washington, DC: American Psychiatric Publishing, Inc; 2001.
49. Mezzich JE, Caracci G, Fabrega H Jr, et al. Cultural formulation guidelines. *Transcult Psychiatry.* 2009 Sep; 46(3):383–405. [PubMed: 19837778]
50. Martínez LC. DSM-IV-TR cultural formulation of psychiatric cases: two proposals for clinicians. *Transcult Psychiatry.* 2009 Sep; 46(3):506–523. [PubMed: 19837784]
51. Aggarwal NK. Hybridity and intersubjectivity in the clinical encounter: impact on the cultural formulation. *Transcult Psychiatry.* 2012 Feb; 49(1):121–139. [PubMed: 22218399]
52. Aggarwal NK. Adapting the cultural formulation for clinical assessments in forensic psychiatry. *J Am Acad Psychiatry Law.* 2012; 40(1):113–118. [PubMed: 22396348]
53. Aggarwal NK. Cultural psychiatry, medical anthropology, and the DSM-5 field trials. *Med Anthropol.* 2013 in press.
54. Aggarwal NK, Nicasio AV, DeSilva R, et al. Barriers to implementing the DSM-5 Cultural Formulation Interview: a qualitative study. *Cult Med Psychiatry.* 2013 in press.
55. Fraser MW, Galinsky MJ. Steps in intervention research: designing and developing social programs. *Research on Social Work Practice.* 2010; 20(5):459–466.

56. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. fifth edition. Arlington: American Psychiatric Publishing, Inc; 2013.
57. Bhui K, Warfa N, Edonya P, et al. Cultural competence in mental health care: a review of model evaluations. *BMC Health Serv Res.* 2007 Jan 7.;15. [PubMed: 17266765]
58. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from Stage I. *Clinical Psychology: Science and Practice.* 2001; 8(2):133–142.
59. Gearing RE, El-Bassel N, Ghesquiere A, et al. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clin Psychol Rev.* 2011 Feb; 31(1):79–88. [PubMed: 21130938]
60. Kazdin AE. Comparative outcome studies of psychotherapy: methodological issues and strategies. *J Consult Clin Psychol.* 1986 Feb; 54(1):95–105. [PubMed: 3958306]
61. Barber JP, Mercer D, Krakauer I, et al. Development of an adherence/competence rating scale for individual drug counseling. *Drug Alcohol Depend.* 1996 Dec 11;43(3):125–132. [PubMed: 9023068]
62. O'Malley SS, Foley SH, Rounsaville BJ, et al. Therapist competence and patient outcome in interpersonal psychotherapy of depression. *J Consult Clin Psychol.* 1988 Aug; 56(4):496–501. [PubMed: 3198805]
63. Barber JP, Crits-Christoph P. Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: a preliminary report. *Psychother Res.* 1996 Nov; 6(2):81–94. [PubMed: 22242608]
64. Suzuki Y, Takahashi T, Nagamine M, et al. Comparison of psychiatrists' views on classification of mental disorders in four East Asian countries/area. *Asian J Psychiatr.* 2010 Mar; 3(1):20–25. [PubMed: 23051133]
65. Reed GM, Mendonça Correia J, Esparza P, et al. The WPA-WHO Global Survey of Psychiatrists' Attitudes Towards Mental Disorders Classification. *World Psychiatry.* 2011 Jun; 10(2):118–131. [PubMed: 21633689]
66. Carroll KM, Nich C, Sifry RL, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug Alcohol Depend.* 2000 Jan 1; 57(3): 225–238. [PubMed: 10661673]
67. Waltz J, Addis ME, Koerner K, et al. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *J Consult Clin Psychol.* 1993 Aug; 61(4):620–630. [PubMed: 8370857]
68. Dusenbury L, Brannigan R, Falco M, et al. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res.* 2003 Apr; 18(2):237–256. [PubMed: 12729182]
69. Nichter M. Idioms of distress: alternatives in the expression of psychosocial distress: a case study from South India. *Cult Med Psychiatry.* 1981 Dec; 5(4):379–408. [PubMed: 7326955]
70. Williams N, Ogden J. The impact of matching the patient's vocabulary: a randomized control trial. *Fam Pract.* 2004 Dec; 21(6):630–635. [PubMed: 15520032]
71. Moyers TB, Martin T, Manuel JK, et al. Assessing competence in the use of motivational interviewing. *J Subst Abuse Treat.* 2005 Jan; 28(1):19–26. [PubMed: 15723728]
72. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychological Assessment.* 1995 Sep; 7(3):309–319.
73. Schoenwald SK, Garland AF, Chapman JE, et al. Toward the effective and efficient measurement of implementation fidelity. *Adm Policy Ment Health.* 2011 Jan; 38(1):32–43. [PubMed: 20957425]
74. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005 Nov; 15(9):1277–1288. [PubMed: 16204405]
75. Cohen J. A coefficient agreement for nominal scales. *Educ Psychol Meas.* 1960; 20(1):37–46.
76. Spearman C. The proof and measurement of association between two things. *Amer J Psychol.* 1904 Jan; 15(1):72–101.
77. Grove WM, Andreasen NC, McDonald-Scott P, et al. Reliability studies of psychiatric diagnosis: theory and practice. *Arch Gen Psych.* 1981; 38(4):408–413.

78. Carroll KM, Nuro KF. One size cannot fit all: a stage model for psychotherapy manual development. *Clin Psychol Sci Prac.* 2002 Dec; 9(4):396–406.
79. Renzaho AM, Romios P, Crock C, et al. The effectiveness of cultural competence programs in ethnic minority patient-centered health care--a systematic review of the literature. *Int J Qual Health Care.* 2013 Epub ahead of print.

Table 1

Sample characteristics

	Total (n=32)
<i>Patients</i>	
Mean age (SD)	53.50 (13.69)
Female	22
Race/ethnicity	
Non-Hispanic White	4
African American	5
Hispanic/Latino	22
Missing	1
Employed (at least part-time)	5
Primary diagnosis Post-CFI	
Depression	14
Bipolar	7
Schizophrenia	8
Other	3
Other specified	2 (OCD & Psychosis NOS)
	Total (n=7)
<i>Clinicians</i>	
Mean age (SD)	38.86 (9.92)
Female	5
Race/ethnicity	
Non-Hispanic White	2
Hispanic/Latino	3
Other	2
Profession	
Psychiatrist	3
Psychologist	1
Social worker	2
Other	1

Table 2

Measures of concordance between raters and percent of positive responses for each CFI question for two fidelity dimensions (n=23)

CFI-FI question based on CFI topic: "Did the clinician ask a CFI question about ..."	Adherence to method		Patient responsiveness	
	Asked by clinicians	Rater concordance	Relevant Patient response	Rater concordance
1. The patient's presenting problem?"	23 (100%)	23 (100%)	22 (96%)	20 (87%)
2. The patient's description of the problem to members of the patient's social network?"	22 (96%)	22 (96%)	21 (91%)	20 (87%)
3. The severity of the problem?" (The patient may not be experiencing any severity at all and this should still be coded.)	22 (96%)	17 (74%)	23 (100%)	17 (74%)
4. The causes of the problem from the patient's perspective?"	23 (100%)	23 (100%)	22 (96%)	23 (100%)
5. What makes the problem better?"	22 (96%)	22 (96%)	23 (100%)	22 (96%)
6. What makes the problem worse?"	23 (100%)	23 (100%)	23 (100%)	23 (100%)
7. How identity relates to the patient's problem?"	20 (87%)	19 (83%)	23 (100%)	19 (83%)
8. Self-coping?"	22 (96%)	23 (100%)	23 (100%)	23 (100%)
9. Past help seeking?"	23 (100%)	23 (100%)	23 (100%)	22 (96%)
10. Barriers to care?"	22 (96%)	23 (100%)	23 (100%)	23 (100%)
11. The patient's current treatment preferences?"	23 (100%)	23 (100%)	23 (100%)	23 (100%)
12. How the clinician's identity may cause barriers to care?"	22 (96%)	23 (100%)	23 (100%)	23 (100%)

Table 3

Frequencies of sum scores for adherence to method and patient responsiveness

Fidelity subscale	Fidelity score	Overall Frequency (Percent)	Frequency by language (%)			Frequency by clinician type (%)	
			English (n=11)	Spanish (n=12)	MD (n=9)	Non-MD (n=14)	
Asked (Adherence to method)	10	2 (8.7)	1 (9.1)	1 (8.3)	1 (11.1)	1 (7.1)	
	11	5 (21.7)	3 (27.3)	2 (16.7)	2 (22.2)	3 (21.4)	
	12	16 (69.6)	7 (63.6)	9 (75.0)	6(66.7)	10 (71.4)	
Responded (Patient responsiveness)	10	2 (8.7)	1 (9.1)	1 (8.3)	1 (11.1)	1 (7.1)	
	11	5 (21.7)	3 (27.3)	2 (16.7)	2 (22.2)	3 (21.4)	
	12	16 (69.6)	7 (63.6)	9 (75.0)	6(66.7)	10 (71.4)	
Good response (Patient responsiveness)	9	1 (4.4)	1 (9.1)	0	1 (11.1)	0	
	10	3 (13.0)	1 (9.1)	2 (16.7)	1 (11.1)	2 (14.3)	
	11	4 (17.4)	3 (27.3)	1 (8.3)	2 (22.2)	2 (14.3)	
	12	15 (65.2)	6 (54.6)	9 (75.0)	5 (55.6)	10 (71.4)	

“Fidelity score” refers to the score for this subscale on the CFI-FL.

Table 4

Inter-rater reliability and fidelity scores for clinician competence and intervention distinctness items (n=23)

Question	Percent concordant	% responded		
		0	1	2
<i>Clinician competence</i>				
Empathy: Did the clinician paraphrase or name the patient's emotional state?	65.2%	8.7%	26.1%	65.2%
Patient centeredness: Did the clinician maintain a non-judgmental attitude (not arguing, confronting, or correcting the patient)?	95.7%	0	4.3%	95.7%
Clarification: Did the clinician ask follow-up questions to understand unclear patient responses?	100%	0	0	100%
Word matching: Did the clinician use the patient's preferred illness term whenever the CFI question stem included the term "[PROBLEM]"?	52.2%	13%	8.7%	78.3%
Illness narration: Did the clinician's interactions help the patient construct and explore a narrative account of illness or did the clinician seem to rush through the CFI?	95.7%	0	0	100%
<i>Intervention distinctness</i>				
Drift: Did the clinician ask about topics during the CFI session that typically belong to the standard clinical interview (history of present illness, current medications, detailed psychiatric or medical history, family history, social history, mini-mental status examination)?	82.6%	91.3%	8.7%	0
Order: Did the clinician ask about all topics in order as reflected in the CFI clinician guidelines?	100%	0	0	100%

Table 5

Correlation among all CFI-FI variables measuring intervention fidelity (n=23)

Construct	Adherence to method	Patient Responsiveness		Clinician Competence			Intervention Distinctness	Dosage
		Asked Total Score	Responded Total Score	Good Response Total Score	Empathy	Patient Centeredness		
Adherence to method	Asked Total Score	1.00	r = 1.00 p < .0001	r = .92 p < .0001	r = -.15 p = .50	r = .28 p = .20	r = .20 p = .36	r = .00 p = 1.00
	Responded Total Score		1.00	r = .92 p < .0001	r = -.15 p = .50	r = .28 p = .20	r = .20 p = .36	r = 0 p = 1.00
Patient Responsiveness	Good Response Total Score	r = .92 p < .0001	r = .92 p < .0001	1.00	r = .04 p = .87	r = .34 p = .11	r = .22 p = .31	r = -.11 p = .64
	Empathy	r = -.15 p = .50	r = -.15 p = .50		1.00	r = .25 p = .25	r = .22 p = .31	r = .18 p = .41
Clinician competence	Patient Centeredness	r = .28 p = .20	r = .28 p = .20	r = .34 p = .11	r = .25 p = .25	1.00	r = .07 p = .77	r = -.09 p = .70
	Word Matching	r = -.16 p = .46	r = -.16 p = .46	r = .06 p = .78	r = .22 p = .30	r = -.11 p = .61	r = -.24 p = .27	r = -.53 p = .01
	Drift	r = .20 p = .36	r = .20 p = .36	r = .22 p = .31	r = .22 p = .31	1.00	1.00	r = .37 p = .09
Dosage	Time	r = 0 p = 1.00	r = 0 p = 1.00	r = -.11 p = .64	r = .18 p = .41	r = -.09 p = .70	r = .37 p = .09	1.00