



Published in final edited form as:

J Subst Abuse Treat. 2015 August ; 55: 52–57. doi:10.1016/j.jsat.2015.02.006.

Adapting the Helpful Responses Questionnaire to assess communication skills involved in delivering contingency management: Preliminary psychometrics

Bryan Hartzler

Abstract

A paper/pencil instrument, adapted from Miller and colleagues' (1991) Helpful Responses Questionnaire (HRQ), was developed to assess clinician skill with core communicative aspects involved in delivering contingency management (CM). The instrument presents a single vignette consisting of six points of client dialogue to which respondents write 'what they would say next.' In the context of an implementation/effectiveness hybrid trial, 19 staff clinicians at an opiate treatment program completed serial training outcome assessments before, following, and three months after CM training. Assessments included this adaptation of the HRQ, a multiple-choice CM knowledge test, and a recorded standardized patient encounter scored for CM skillfulness. Study results reveal promising psychometric properties for the instrument, including strong scoring reliability, internal consistency, concurrent and predictive validity, test-retest reliability and sensitivity to training effects. These preliminary findings suggest the instrument is a viable, practical method to assess clinician skill in communicative aspects of CM delivery.

1. Introduction

The dissemination of empirically-supported behavior therapies to routine clinical care remains a poignant challenge, for therapy purveyors as well as the clinic directors and direct-care staff who would be eventual agents of therapy implementation (Carroll, 2012). A promising trend in the past decade has been greater awareness of behavior therapies in community settings, fueled by large-scale efforts of the Substance Abuse Mental Health Services Administration (SAMHSA) and National Institute on Drug Abuse (NIDA). Prominent undertakings include SAMHSA's maintenance of a National Registry of Evidence Based Programs and Practices (www.nrepp.samhsa.gov) and regional Addiction Technology Transfer Centers, NIDA's creation of a Clinical Trials Network to spur multisite effectiveness research (Hanson, Leshner, & Tai, 2002), and a joint effort to develop clinician-friendly 'blending products' (Martino et al., 2010). Even so, prior research suggests adoption of even widely-promoted empirically-supported treatments occurs among

© 2015 Published by Elsevier Inc.

Corresponding author: Bryan Hartzler, Ph.D., Alcohol & Drug Abuse Institute, University of Washington, Box 354805, 1107 NE 45th Street, Suite 120, Seattle, WA 98105-4631, U.S.A., 206.543.8369, 206.543.5473 (fax), hartzb@u.washington.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

just a minority of community addiction treatment settings (Knudsen, Abraham, & Roman, 2011; Roman, Abraham, Rothrauff, & Knudsen, 2010).

Beyond treatment community awareness, the challenge of behavior therapy dissemination includes barriers beyond the initial adoption decision. Many barriers implicate quality assurance that a therapy is skillfully delivered (McHugh & Barlow, 2010). Questions remain about what training and supervision methods are sufficient to develop and maintain skillful delivery of empirically-supported behavior therapies (Beidas & Kendall, 2010; Herschell, Kolko, Baumann, & Davis, 2010). A linked concern is limited feasibility of the observational systems used in controlled treatment trials to measure the quality of how clinicians deliver a therapy. These systems, typically reliant on time-consuming direct observation or review of recorded clinical sessions paired with structured skill ratings (Baer et al., 2007), are unwieldy for under-resourced treatment settings. Further, the intrusive nature of observational processes presents logistical and philosophical challenges for community-based clinicians and their clientele. Accordingly, there is need to develop and validate practical skill assessment methods, and a key component of such methods is that they prompt clinicians to rehearse therapy delivery skills (Beidas, Cross, & Dorsey, 2014). In many therapies, such therapy delivery skills reflect how clinicians respond verbally to clients. Several validated skill assessment methods simulate this using standardized clinical stimuli—in the form a live standardized patient (Imel et al., 2014; Stimmel, Cohen, Fallar, & Smith, 2006), pre-recorded video clips (Baer et al., 2012; Rosengren, Hartzler, Baer, Wells, & Dunn, 2008), or written vignettes (Miller et al., 1991). A shared attribute of these methods is the opportunity they provide to measure targeted aspects of clinician verbal behavior when the clinician is given a consequence-free opportunity to rehearse communicative skills.

One empirically-supported behavior therapy for treating substance abusers for which options for clinician skill assessment are limited is contingency management (CM), which encompasses a family of related behavioral reinforcement approaches. Petry (2012) notes as core tenets of CM methods that: 1) a focal, desired patient behavior be closely monitored, 2) a tangible, positive reinforcer be provided when the behavior occurs, and 3) reinforcers be withheld when the behavior does not occur. Meta-analyses document reliable therapeutic effects of CM on substance abusers (Dutra et al., 2008; Griffith, Rowan-Szal, Roark, & Simpson, 2000; Lussier, Heil, Mongeon, Badger, & Higgins, 2006; Prendergast, Podus, Finney, Greenwell, & Roll, 2006). And like many empirically-supported behavior therapies, a validated observational measurement system does exist whereby clinician performance in a clinical session may be rated (Petry, Alessi, Ledgerwood, & Sierra, 2010). While well-suited for use in controlled treatment trials, such observational systems are not in wide community use given a mismatch between the personnel resources they require and that which are available in most treatment settings. Thus, design of a practical, skill assessment method would fill an important gap.

Research on training community treatment professionals in CM and its impact on their consequent implementation experience is mixed. Results of a nationwide training effort with VA clinic leaders suggests multi-day workshop exposure promotes conducive beliefs about CM (Rash, DePhillips, McKay, Drapkin, & Petry, 2013). Further, many of the

corresponding VA clinics reportedly attempted to institute CM in some form when provided initial funding support (Petry, DePhillips, Rash, Drapkin, & McKay, 2014). However, this national VA initiative did not assess the nature, quality, or sustainment of those implementation efforts, and extant literature contains multiple examples of well-intentioned attempts to implement CM undermined or discontinued after encountering of logistical or procedural problems (Tuten, Svikis, Keyser-Marcus, O'Grady, & Jones, 2012; Walker et al., 2010). Critical to successful implementation of CM is the capacity of direct-care clinicians to demonstrate core communication skills they would be expected to display during delivery of contemporary CM interventions. Thus, a practical method for assessing these core communication skills is needed.

A recent CM implementation/effectiveness trial at a community opiate treatment program (Hartzler, Jackson, Jones, Beadnell, & Calsyn, 2014), which principally evaluated impacts of training among staff clinicians, offered opportunity to develop and evaluate psychometric properties of a pencil/paper skill assessment instrument. This instrument, adapted from the Helpful Responses Questionnaire [HRQ; (Miller et al., 1991)] originally developed to assess clinician skill in communicating empathy, sought to provide a practical way to assess core communicative skills involved in delivery of contemporary CM interventions. The trial included serial training outcome assessments for participating clinicians—occurring prior to, immediately following, and three months after training in a quasi-experimental design that also accounted for assessment reactivity. In each training outcome assessment, a new version of the HRQ adapted for contingency management (HRQ-CM) was administered alongside an existing CM knowledge test and standardized patient interview scored with a validated observational CM measurement system. Upon completing training, clinicians had opportunity to implement a CM intervention with targeted clients on their caseload on a provisional basis for 90 days. Herein, preliminary psychometrics of the HRQ-CM are reported including scoring reliability, internal consistency, test-retest reliability, concurrent and predictive validity, and sensitivity to training effects. A data-informed approach is also taken in proposing a provisional competency benchmark.

2. Materials and method

2.1 Parent Trial Design

This was an implementation/effectiveness hybrid trial design (Curran, Bauer, Mittman, Pyne, & Stetler, 2012), for which implementation and clinical effectiveness outcomes have been comprehensively reported (Hartzler et al., 2014). The trial included: 1) recruitment of interested staff clinicians to participate in CM training and implement a contextualized CM intervention with targeted clients on their caseload for a 90-day period, and 2) serial completion of training outcome assessments prior to, following, and three months after training. To account for potential assessment reactivity (given absence of no-training/waitlist control condition), the clinicians were randomly-assigned to a single baseline assessment condition completed a week prior to training or a repeated baseline assessment condition with measures completed two weeks prior to training and repeated a week later. This quasi-experimental trial design feature enabled analysis of the HRQ-CM's test-retest reliability in a subsample of ten clinicians.

2.2 Collaborating Treatment Setting

The collaborating setting is a private, non-profit opiate treatment program located in an urban area of a large U.S. city. It maintains a census of 1000 patients who receive agonist medication, individual/group counseling, and monthly drug screen urinalysis (UA). The clinic is affiliated with the NIDA Clinical Trials Network, and had previously participated in multisite trials of alternative treatment approaches as well as other federally-funded research.

2.3 Contextualized CM Intervention

A full description of this CM intervention and its collaborative design process is available (Hartzler et al., 2014). Briefly, and at the clinic director's specification, the CM intervention featured: 1) a target population of patients in their initial 90 days of clinic services, 2) a target behavior of individual counseling visit attendance, 3) \$5 gift cards and take-home medication doses as reinforcers, and 4) a 'point-based' reinforcement system. An investigator-derived reinforcement schedule integrated priming and escalation/reset features to enhance clinical impacts. In a 90-day clinic implementation period, trained clinicians delivered this intervention on a trial basis with eligible clients on their caseload—monitoring the target behavior, tracking earned points, and delivering reinforcers amidst usual care in weekly counseling visits. To aid clinic tracking, the electronic medical record system was adapted to include documentation of patient point totals (and any reinforcers provided) in individual counseling visit notes.

2.4 Adapted Helpful Responses Questionnaire for Contingency Management (HRQ-CM)

The HRQ-CM was designed as a paper/pencil method to assess communication skills involved in delivery of CM interventions. It sought to capitalize on the structural appeal and practicality of the original and previously-validated HRQ instrument developed by Miller and colleagues (1991), who intended it to be a practical measure of clinician skill in communicating empathy. Corresponding observational measurement systems (Truax & Carkhuff, 1967) were thought to be a poor match for many in the treatment community, given their cumbersome and resource-intensive nature. Miller and colleagues also suspected poor correlation between clinician self-reports of clinical practice behavior and objective behavioral ratings by 3rd-parties, a hypothesis borne out in later research (Baer et al., 2004; Miller & Mount, 2001; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004). The original HRQ was structured in vignette form with six discrete patient scenarios presented, to which the respondent was to write "the next thing you would say" to each patient. In its original validation, the HRQ evidenced strong scoring reliability by independent raters ($r=.93$), internal consistency (Cronbach $\alpha = .92$), and sensitivity to the effects of training (Miller et al., 1991). Consequently, the HRQ offered a compelling template from which the current work sought methodological adaptation to tap clinician skill with core communicative aspects of CM delivery.

The HRQ-CM was designed to assess a set of six core communication skills involved in CM delivery, which may all occur in a given counseling visit. Consequently, the basic vignette structure of the instrument was maintained but with item content contextually linked in one patient scenario as if occurring in a single visit. Communication skills targeted for

measurement were as specified in an available version of the Contingency Management Competence Scale [CMCS; (Petry & Ledgerwood, 2010)], an observational coding measure with strong prior psychometric validation (Petry et al., 2010). The author drafted an initial set of vignette stimuli for the six HRQ-CM items, and later revised them in accord with feedback from two former addiction treatment program directors. Table 1 lists the content of these eventual HRQ-CM items, and the communication skills they tap via six-point scales (0 = Absent, 5 = Excellent). This format produces an HRQ-CM summary score ranging from 0–30. Administration time for the HRQ-CM during this trial was consistently 5–10 minutes.

2.5 Procedures

All procedures were approved by university IRB, and those detailing the 16-hour staff training process and 90-day provisional implementation of this CM intervention are available elsewhere (Hartzler et al., 2014). Staff clinicians were recruited at an on-site presentation where voluntary participation was emphasized by the investigator and program director. Interested clinicians provided informed consent and then notified of their baseline assessment condition. All other study activities were uniform for clinicians, and included subsequent assessments one week after training and after conclusion of the 90-day provisional implementation period.

The initial baseline assessment elicited the demography and professional background of clinicians. In addition to the HRQ-CM, all assessments included an 18-item multiple-choice CM Knowledge Test (Petry & Stitzer, 2002) that measured conceptual knowledge of CM principles and practices, for which adequate internal consistency ($\alpha = .76$) led to computation of a summary score. All assessments also included a Standardized Patient (SP) Encounter, wherein an actor experienced in university-based psychotherapy education projects was hired to present at clinicians' offices and enact a 20-minute, audio-recorded encounter with a newly-enrolled patient. Initially drafted by the author, this SP character/scenario incorporated input from two former addiction treatment program directors. The SP received eight hours of training, with a conceptual orientation to behavioral reinforcement principles and practices followed by extensive practice of SP character portrayal via role-plays with the author. The SP character presentation was consistent (not progressive) across the serial training outcome assessments, and this maximized opportunity to assess temporal changes in focal clinician communication skills.

Distinct pairs of previously CM-naïve raters scored the HRQ-CMs and CMCS-rated SP encounters. For HRQ-CMs, raters were trained via a two-hour conceptual orientation to CM principles and practices followed by two hours of applied practice in which they scored sample HRQ-CM responses created by the author. After training, each rater—blinded to the timing of assessment—independently scored the 64 HRQ-CMs collected in the trial over the course of eight hours, such that instrument scoring required 7–8 minutes apiece. For SP encounters, raters were trained via the same two-hour conceptual orientation to CM principles and practices followed by four hours of applied practice scoring sample SP encounters in which the author intentionally portrayed clinicians of varying skillfulness. After training, each rater scored the trial's 64 SP encounters (blind to timing of assessment) using the CMCS (Petry & Ledgerwood, 2010). This required 28–30 hours per rater, such

that scoring of an SP encounter took on average 26–28 minutes. The excellent inter-rater reliability observed across six CMCS skill domains (ICCs=.77–.89) enabled computation of a CMCS summary score used in later analyses.

2.6 Participants

Participants were 19 staff clinicians active in delivery of program services. The mean staff age was 59.32 years (S.D.=12.73). A large majority (89%) were female. Hispanic ethnicity was identified by 5%, and distribution of race was 79% Caucasian, 16% Multi-racial, and 5% Native American. In terms of education, 58% had masters, 26% bachelors, and 16% associate degrees. Many were long-standing employees, as mean clinic tenure was 12.24 years (S.D.= 9.72). With respect to prior CM exposure, 11% had attended a presentation, 31% had reviewed published or online works, 27% noted both types of exposure, and 31% reported no exposure.

2.7 Data Analytic Strategy

Analysis of HRQ-CM psychometrics was a multi-step process. A 1st step focused on scoring reliability, evaluated at an item-level using all HRQ-CMs completed in the trial. Item-specific intra-class correlations (ICCs) assessed rater agreement, and were interpreted relative to psychometric standards (Cicchetti, 1994). A 2nd step targeted internal consistency, using initial baseline assessment data from all 19 participating clinicians. A scale reliability analysis obtained a Cronbach alpha, supplemented by item-level descriptive statistics and item-scale correlations. A 3rd step examined test-retest reliability, using paired baseline assessments for the ten clinicians in the repeated baseline assessment condition. A repeated-measures analysis of variance (RM-ANOVA) assessed temporal change in HRQ-CM summary scores, and Cohen's *d* for dependent measures was computed and interpreted per psychometric standards (Cohen, 1988). A 4th step determined concurrent validity, using initial baseline assessment data from all 19 clinicians with Pearson correlations computed among summary scores of the HRQ-CM, CM Knowledge Test, and CMCS-rated SP encounter. A 5th step assessed instrument sensitivity to effects of training, using initial baseline and post-training assessment data from all 19 clinicians. A RM-ANOVA examined temporal change in HRQ-CM summary scores attributable to training, with a Cohen's *d* for dependent measures effect size computed and interpreted per psychometric standards and corresponding effect sizes for the CM Knowledge Test and CMCS-rated SP encounters. A 6th step examined predictive validity, via Pearson correlation of clinicians' post-training HRQ-CM summary score and a caseload-aggregated clinical effectiveness index (reflecting counseling visit attendance—the client behavior targeted for reinforcement) from the parent trial's 90-day period of implementation. This index was computed as the percentage of scheduled counseling visits attended by the subset of clients on each clinician's caseload eligible for the CM intervention.

A final step was to explore competency benchmarking for the instrument, guided by the distribution of HRQ-CM summary scores assessed at trial conclusion. This was informed by previously-reported findings (Hartzler et al., 2014) suggesting this clinician sample to be a capable, experienced, and ecologically-valid reference group. Findings included: 1) robust, durable CM training impacts on clinician skills in CMCS-rated SP encounters completed

before, after, and 90 days following training, 2) impacts of the resulting clinician implementation ($d = .46-.53$) exceeded the mean effect size reported in a published meta-analysis of CM efficacy trials (Prendergast et al., 2006), 3) an absence of clinician effects in these analyses, suggesting consistency in both quality of implementation and clinical impact, and 4) each clinician had exceeded a competency threshold suggested by CMCS developers when completing the SP encounter at trial conclusion. Consequently, the distribution of HRQ-CM summary scores collected at trial conclusion was reviewed, with central tendency among this clinician reference group explored to inform a threshold serving as a provisional HRQ-CM competency benchmark.

3. Results

3.1 Scoring Reliability

Item-specific ICCs were: 1) informing patient of earned reinforcers, .84; 2) informing patient of prospective reinforcers, .93; 3) delivering earned reinforcers, .84; 4) assessing patient interest in reinforcers, .89; 5) praising patient efforts to earn reinforcers, .77; and 6) linking the reinforced behavior to patient abstinence and other treatment goals, .82. Further, the ICC for the HRQ-CM summary score was .96. Notably, all of these ICC values fall in what Cicchetti (1994) proposes as 'excellent' range for inter-rater reliability. As a result, the dataset for subsequent analyses consisted of the mean of HRQ-CM ratings provided by the two independent raters.

3.2 Internal Consistency

Based on initial baseline administrations, internal consistency of the six HRQ-CM items was reasonable (Cronbach $\alpha = .74$) with no suggestion of improvement via item removal or their reconstitution as subscales. Table 2 presents corresponding item-level descriptive statistics and item-scale correlations. Item-scale correlations show positive association of all six HRQ items with its summary score ($r = .36-.62$), suggesting respective contribution of all six items to the overall scale. Accordingly, subsequent analyses utilized the HRQ-CM summary score as a unitary reflection of practitioners' skill with the targeted communicated aspects of CM delivery.

3.3 Test-Retest Reliability

The RM-ANOVA did not detect significant temporal change in the HRQ-CM summary scores between its initial and repeated baseline administrations ($p = .75$). In fact, the mean score of initial HRQ-CMs was *slightly higher* ($M = 13.01$; $SD = 5.79$) than for those repeated a week later ($M = 12.91$, $SD = 5.72$). The effect size for this negative trend was nominal ($d = .03$), and subsequent analyses of HRQ-CM sensitivity to training effects were expanded to the full sample.

3.4 Concurrent Validity

Concurrent validity was evaluated via examination of the baseline association of the HRQ-CM summary score with those of the CM Knowledge Test and CMCS-rated SP encounter. Broadly, analyses identified positive associations among all three measures. Though the positive association ($r = .41$, $p < .08$) between the CM Knowledge Test and CMCS-rated SP

encounter only trended toward statistical significance, the HRQ-CM was significantly correlated with both the CM Knowledge Test ($r = .55$) and CMCS-rated SP encounter ($r = .53$).

3.5 Sensitivity to Training Effects

A RM-ANOVA revealed significant pre-to-post training temporal increase in HRQ-CM summary scores, $F(1,18) = 44.03, p < .001$. Specifically, a mean improvement of more than eight points (pre-training $M=12.92, SD=5.59$, post-training $M=21.39, SD=3.10$) was evidenced by this clinician sample. The corresponding pre-to-post-training effect size for the HRQ-CM summary score ($d = 1.52$) is juxtaposed between those previously reported by Hartzler and colleagues (2014) for the CM Knowledge Test ($d = .1.10$) and CMCS-rated SP encounters ($d = 2.09$).

3.6 Predictive Validity

Initial examination of the involved variable distributions confirmed reasonable range for this small clinician sample. Post-training HRQ summary scores ranged from 17.50 to 27.50, with a median (21.00) proximal to the sample mean (21.39). The caseload-aggregated clinical effectiveness index, computed as a percentage given counseling visit attendance as the client behavior targeted for reinforcement, ranged from 67–100% ($M=83.21, Md=86.00, S.D.=11.47$). Post-training HRQ summary scores showed evidence of predictive validity for client outcomes, as their positive association with this caseload-aggregated clinical effectiveness index reached statistical significance ($r = .52, p < .05$).

3.7 Provisional Competency Benchmarking

Distributional properties reflecting the central tendency of summary scores from HRQ-CMs completed by this clinician sample at trial conclusion were of relevant interest, given the set of previously-reported findings from the parent trial (Hartzler et al., 2014) suggesting it to be a suitable reference group. The mean and median of this HRQ-CM summary score distribution were closely-aligned ($M=20.57, Md=20.75$), with upper and lower quartile quartiles of 18.63 and 23.00. The intent of competency benchmarking for the HRQ-CM is primarily to guide those in the treatment community who may use this instrument to gauge clinicians' skill in core communicative aspects of delivering CM interventions, and thus the use of an integer to mark this threshold holds practical value. Accordingly, an HRQ-CM summary score of 20 was chosen as a provisional competency benchmark for this instrument.

4. Discussion

The addiction treatment field continues to seek dissemination of empirically-supported behavior therapies like CM, and this relies in part on availability of instruments that assess the communication skills of clinicians in a valid, practical manner. An implementation/effectiveness hybrid trial (Hartzler et al., 2014) presented opportunity to develop and evaluate the HRQ-CM, a pencil/paper instrument adapted from Miller's (1991) Helpful Responses Questionnaire to assess clinician skill in communicative aspects of CM delivery. Preliminary psychometric properties revealed of the HRQ-CM in the current work are

promising, and include: 1) scoring reliability among independent raters, 2) internal consistency, eventuating in a unitary summary score, 3) one-week test-retest reliability, 4) concurrent validity, with significant association to knowledge- and skills-based instruments for CM, 5) sensitivity to impacts of clinician training commensurate with that previously-reported for other knowledge- and skills-based measures, and 6) predictive validity for targeted clinical outcomes of clients with whom a CM intervention was subsequently implemented. Given previously-reported evidence of strong and consistent CM delivery skills among this clinician sample and clinical effectiveness of their corresponding implementation, a threshold reflecting central tendency of their HRQ-CM summary scores at trial conclusion was chosen to serve as a provisional competency benchmark for this instrument.

The question of how to assess communication skills of addiction treatment professionals implicates a host of considerations, for which measurement validity understandably is at the forefront. Many paper/pencil assessment methods gather clinician-reports of practice behavior, which—while efficient with respect to time and resources—are subject to demand characteristics that limit their validity. Long-standing recognition of this by treatment researchers has prompted design of a variety of tailored skills-based measures, including Miller and colleagues' (1991) original HRQ as well as a host of comprehensive, therapy-specific coding systems. Such coding systems are used in controlled treatment trials to verify therapy delivery as intended, and are appropriately regarded a 'gold-standard' for this purpose (Baer et al., 2007). Therapy purveyors who promote such coding systems as means for community settings to monitor clinical practice behavior should consider whether this matches the resources available in those settings. Further, they may consider an often-overlooked threat to validity associated with use of such systems. Because behavior therapies involve dyadic clinician-patient interaction, variability in ratings of corresponding clinician behavior will necessarily be influenced by factors related to the clinician, patient, clinician-patient relationship, and sources of error (Kenny, Kashy, Cook, & Simpson, 2006). Given the diversity of clinical case mixes in most treatment settings and selection biases likely to intrude when staff are asked to make work-samples available for external review, there are reasons to question the validity of data gathered via observational methods. Indeed, recent work documents derivation of more reliable, unbiased estimates of clinician competence via skill assessment methods involving standardized clinical stimuli (Imel et al., 2014). In its elicitation of written clinician responses to a standardized patient vignette, the HRQ-CM offers a simple method wherein core communications skills involved in delivering CM interventions may be assessed with lesser concern for this threat to validity.

Beyond instrument psychometric properties and conceptual discussion of measurement validity, treatment community professionals may appreciate some discussion of the practicality of the HRQ-CM relative to observational therapy coding systems. Relative to the CMCS-rated SP encounters included in this trial, the HRQ-CM offered a briefer and less resource-intensive administration for the participating clinicians as well as less time-consuming processes of rater training and eventual scoring. The paper/pencil nature of this HRQ-CM method would allow treatment settings to measure targeted clinician communication skills in a simple, practical manner that avoids the logistical hurdles posed by traditional observational methods (e.g., patient identification and consenting, staff

schedule coordination, recording equipment set-up, security of clinical care data voiced in clinician-patient interactions). Further, the current study relied on previously CM-naïve individuals to serve as raters—who completed conceptual and applied rater training activities and scoring of 64 HRQ-CMs in a span of 12 hours. In addition to the economy of time, HRQ-CM scoring did not require other resources typically involved with observational systems (e.g., real-time session observation or review by supervisor; availability of recording playback equipment, clinical care data storage). In these ways, the HRQ-CM may hold appeal for addiction treatment professionals working in resource-challenged settings.

This work contains a number of caveats. Those most prominent are the involvement of a single community treatment setting and voluntary participation of nineteen of its staff clinicians. The size of this clinician sample is admittedly small; thus, preliminary psychometric properties reported herein clearly await replication with larger, more diverse clinician samples. A related caveat concerns possible selection bias, given the voluntary study participation of these staff clinicians. That study participation rested on a willingness to complete serial training outcome assessments, attend CM training sessions, and subsequently deliver a tailored CM intervention to patients during a provisional implementation period. While the vast majority of available staff clinicians chose to participate, a subset did not. Thus, participating clinicians may have had greater interest, motivation, or adoption readiness for CM than their peers. At a program-level, this was a CTN-affiliate opiate treatment program which had previously participated in research studies concerning empirically-supported behavior therapies. Thus, care should be taken with any generalization of current findings to the broader treatment community. A further caveat concerns the breadth of HRQ-CM measurement, which targeted core communication skills in a simple, straightforward manner via vignette presentation of serial points in a single patient visit. While these patient stimuli were sufficient to prompt clinician rehearsal of the intended set of communication skills in writing, the use of such skills in conversation and amidst nuances of routine clinical care that require clinicians to ‘think on their feet’ may pose challenges that were simulated by this instrument. This represents an interesting opportunity for future HRQ-CM adaptation, whereby instrument expansion may enable its capture of a broader set of skills, the same core skills tapped by multiple vignettes of varying difficulty, or both.

Caveats notwithstanding, the current work offers preliminary validation of the HRQ-CM as a practical method to assess clinician skill in core communicative aspects of CM delivery. An attribute of CM that should facilitate its community transportability is a capacity for contextual adaptation (Hartzler, Lash, & Roll, 2012). To that end, others may find it useful to amend details of the patient vignette in the HRQ-CM to better match particular features (i.e., target population, patient behavior, involved reinforcers) of local CM interventions they seek to monitor. This may include expansion of the breadth of HRQ-CM measurement to encompass communicative skills specific to those local CM interventions. Clearly, any such instrument revisions would merit examination of corresponding psychometric properties. In any event, the current work presents preliminary evidence in support of this paper/pencil CM skill measure as a practical alternative to traditional, resource-intensive observational coding methods. Miller and colleagues’ (1991) intent for the original HRQ was to balance the practicality of a paper/pencil instrument with the validity of skills-based

measurement of clinician response to standardized clinical stimuli. The current work extends this methodological approach for valid, practical measurement of clinician skill in a core set of communicative aspects of CM delivery. It is hoped that this adapted instrument assists those in the treatment community in effectively implementing CM interventions, and that the broader measurement approach holds utility for assessment of other psychotherapy skills inherent in empirically-supported behavior therapies.

Acknowledgments

The research reported herein was supported by the National Institute on Drug Abuse of the National Institutes of Health under award number K23DA025678 (*Integrating Behavioral Interventions in Substance Abuse Treatment*, Hartzler PI). The content of this report is solely the responsibility of the author and does not necessarily represent the official views of the National Institute of Health. The author wishes to specifically thank T. Ron Jackson, Molly Carney, Carol Davidson, Michelle Peavey, and Esther Ricardo-Bulis for their support of clinic implementation efforts, as well as Lynette Wright, Brinn E. Jones, and Robert Brooks for their contributions.

References

- Baer JS, Ball SA, Campbell BK, Miele GM, Schoener EP, Tracy K. Training and fidelity monitoring of behavioral interventions in multi-site addictions research. *Drug and Alcohol Dependence*. 2007; 87(2007):107–118. [PubMed: 17023123]
- Baer JS, Carpenter K, Beadnell B, Stoner S, Ingalsbe M, Hartzler B, Drager Z. Computer Assessment of Simulated Patient Interviews (CASPI): Psychometric properties of a web-based system for the assessment of motivational interviewing skills. *Journal of Studies on Alcohol and Drugs*. 2012; 73(1):154–164. [PubMed: 22152673]
- Baer JS, Rosengren DR, Dunn C, Wells E, Ogle R, Hartzler B. An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*. 2004; 73(1):99–106. [PubMed: 14687964]
- Beidas RS, Cross W, Dorsey S. Show me, don't tell me: Behavioral rehearsal as a training and analogue fidelity tool. *Cognitive and Behavioral Practice*. 2014; 21(1):1–11. [PubMed: 25382963]
- Beidas RS, Kendall PC. Training therapists in evidence-based practice: A critical review from a systems-contextual perspective. *Clinical Psychology: Science and Practice*. 2010; 17(1):1–31. [PubMed: 20877441]
- Carroll KM. Dissemination of evidence-based practices: How far we've come, and how much further we've got to go. *Addiction*. 2012; 107(6):1031–1033. [PubMed: 22324509]
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6(4):284–290.
- Cohen, J. *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*. 2012; 50(3):217–226. [PubMed: 22310560]
- Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry*. 2008; 165:179–187. [PubMed: 18198270]
- Griffith JD, Rowan-Szal GA, Roark RR, Simpson DD. Contingency management in outpatient methadone treatment: A meta-analysis. *Drug and Alcohol Dependence*. 2000; 58:55–66. [PubMed: 10669055]
- Hanson GR, Leshner AI, Tai B. Putting drug abuse research to use in real-life settings. *Journal of Substance Abuse Treatment*. 2002; 23:69–70. [PubMed: 12220602]
- Hartzler B, Jackson TR, Jones BE, Beadnell B, Calsyn DA. Disseminating contingency management: Impacts of staff training and implementation at an opiate treatment program. *Journal of Substance Abuse Treatment*. 2014; 46:429–438. [PubMed: 24462242]

- Hartzler B, Lash SJ, Roll JM. Contingency management in substance abuse treatment: A structured review of the evidence for its transportability. *Drug and Alcohol Dependence*. 2012; 122(1–2):1–10. [PubMed: 22153943]
- Herschell AD, Kolko DJ, Baumann BL, Davis AC. The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review*. 2010; 30:448–466. [PubMed: 20304542]
- Imel ZE, Baldwin S, Baer J, Hartzler B, Dunn C, Rosengren DB, Atkins D. Evaluating therapist competence in motivational interviewing by comparing performance with standardized and real patients. *Journal of Consulting and Clinical Psychology*. 2014; 82(3):472–481. [PubMed: 24588405]
- Kenny, DA.; Kashy, DA.; Cook, WL.; Simpson, JA. *Dyadic data analysis*. New York, NY: Guilford Press; 2006.
- Knudsen HK, Abraham AJ, Roman PM. Adoption and implementation of medications in addiction treatment programs. *Journal of Addiction Medicine*. 2011; 5:21–27. [PubMed: 21359109]
- Lussier JP, Heil SH, Mongeon JA, Badger GJ, Higgins ST. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction*. 2006; 101:192–203. [PubMed: 16445548]
- Martino S, Brigham GS, Higgins C, Gallon S, Freese TE, Albright LM, Condon TP. Partnerships and pathways of dissemination: The National Institute on Drug Abuse -Substance Abuse and Mental Health Services Administration Blending Initiative in the Clinical Trials Network. *Journal of Substance Abuse Treatment*. 2010; 38(Supplement 1):S31–S43. [PubMed: 20307793]
- McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*. 2010; 65(2):73–84. [PubMed: 20141263]
- Miller WR, Hedrick KE, Orlofsky D. The Helpful Responses Questionnaire: A procedure for measuring therapeutic empathy. *Journal of Clinical Psychology*. 1991; 47(3):444–448. [PubMed: 2066417]
- Miller WR, Mount KA. A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*. 2001; 29:457–471.
- Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. [Journal]. *Journal of Counseling and Clinical Psychology*. 2004; 72(6):1050–1062.
- Petry, NM. *Contingency Management for substance abuse treatment: A guide to implementing this evidence-based practice*. New York: Routledge; 2012.
- Petry NM, Alessi SM, Ledgerwood DM, Sierra S. Psychometric properties of the Contingency Management Competence Scale. *Drug and Alcohol Dependence*. 2010; 109(1):167–174. [PubMed: 20149950]
- Petry NM, DePhillips D, Rash CJ, Drapkin M, McKay JR. Nationwide dissemination of contingency management: The Veterans Administration Initiative. *The American Journal on Addictions*. 2014; 23(3):205–210. [PubMed: 24724876]
- Petry, NM.; Ledgerwood, DM. *The Contingency Management Competence Scale for Reinforcing Attendance*. Farmington, CT: University of Connecticut Health Center; 2010.
- Petry, NM.; Stitzer, ML. *YUPD Center. Training Series, #6*. West Haven, CT: Yale University; 2002. *Contingency Management: Using motivational incentives to improve drug abuse treatment*.
- Prendergast M, Podus D, Finney JW, Greenwell L, Roll JM. Contingency management for treatment of substance use disorders: A meta-analysis. *Addiction*. 2006; 101:1546–1560. [PubMed: 17034434]
- Rash CJ, DePhillips D, McKay JR, Drapkin M, Petry NM. Training workshops positively impact beliefs about contingency management in a nationwide dissemination effort. *Journal of Substance Abuse Treatment*. 2013; 45:306–312. [PubMed: 23856601]
- Roman PM, Abraham AJ, Rothrauff TC, Knudsen HK. A longitudinal study of organizational formation, innovation adoption, and dissemination activities within the National Drug Abuse

Treatment Clinical Trials Network. *Journal of Substance Abuse Treatment*. 2010; 38(Supplement 1):S44–S52. [PubMed: 20307795]

Rosengren DB, Hartzler B, Baer JS, Wells EA, Dunn C. The video assessment of simulated encounters-revised (VASE-R): Reliability and validity of a revised measure of motivational interviewing skills. *Drug and Alcohol Dependence*. 2008; 97:130–138. [PubMed: 18499356]

Stimmel B, Cohen D, Fallar R, Smith L. The use of standardized patients to assess clinical competence: Does practice make perfect? *Medical Education*. 2006; 40:444–449. [PubMed: 16635124]

Truax, CB.; Carkhuff, RR. *Toward effective counseling and psychotherapy: Training and practice*. Chicago, IL: Aldine; 1967.

Tuten M, Svikis DS, Keyser-Marcus L, O’Grady KE, Jones HE. Lessons learned from a randomized trial of fixed and escalating contingency management schedules in opioid-dependent pregnant women. *American Journal of Drug and Alcohol Abuse*. 2012; 38(4):286–292. [PubMed: 22352784]

Walker R, Rosvall T, Field CA, Allen S, McDonald D, Salim Z, Adinoff B. Disseminating contingency management to increase attendance in two community substance abuse treatment centers: Lessons learned. *Journal of Substance Abuse Treatment*. 2010; 39(3):202–209. [PubMed: 20598838]

Highlights

- Via pencil/paper format, the HRQ-CM assesses clinician skill with CM interventions.
- Preliminary psychometric properties were examined in an implementation/ effectiveness trial.
- Scoring reliability, internal consistency, and concurrent/predictive validity was demonstrated.
- Test-retest reliability and sensitivity to effects of training was also documented.
- A provisional clinician competency benchmark is proposed for the instrument.

Table 1

HRQ-CM Item Content

Item	Targeted CM Delivery Skill
As Joe sits down in your office, you see that he notices a lock-box on the desk entitled 'Starbucks gift-cards.' After you introduce yourself to him, Joe asks "Are <i>those</i> gift cards for clients?"	Informing patient of earned reinforcers
You continue the conversation with Joe, and he mentions an interest in the convenience of obtaining take-home medication doses. He asks you: "How many of these sessions do I need to attend to get a take-home dose?"	Informing patient of prospective reinforcers
Once the point-system has been explained to him, Joe seems torn about which incentive to try for. He says "I'm just not sure whether I should use my points toward a take-home dose or to go to Starbucks."	Assessing patient interest in reinforcers
Later in the conversation, Joe says "The more I think about it, the more I think I do want to go to Starbucks after this counseling session ends. I'm hungry, ya know."	Delivering earned reinforcers
Joe later says "This is a pretty cool incentive program you have here. Out in the real-world, it seems like the only consequences I ever see is someone threatening to punish me for doing something wrong."	Praising patient efforts to earn reinforcers
Later in the session, you and Joe are discussing his long-term goals. Joe says "I really hope things go well for me in this program. I'd like to get off the streets and eventually have my own job again."	Linking reinforced behavior to abstinence and other goals

Notes: Instrument instructions direct the respondent to 'write the next thing you would say to Joe,' who is a 28-year old recently assigned patient who is at your office for his 1st counseling visit; all items scored for their target CM delivery skill, on a 6-point scale (0 = Absent, 5 = Excellent).

Table 2

Item-level Descriptive Statistics and Inter-correlation

	M	(SD)	Corrected item-scale correlation
HRQ-CM #1, Informing patient of earned reinforcers	2.13	(1.68)	.36
HRQ-CM #2, Informing patient of prospective reinforcers	1.40	(1.48)	.48
HRQ-CM #3, Assessing patient interest in reinforcers	3.00	(1.48)	.62
HRQ-CM #4, Delivering earned reinforcers	2.47	(1.25)	.53
HRQ-CM #5, Praising patient efforts to earn reinforcers	2.11	(1.21)	.46
HRQ-CM #6, Linking reinforced behavior to abstinence and other goals	1.82	(1.33)	.36
HRQ-CM Summary Score	12.92	(5.59)	

Notes: Data based on initial baseline assessment with 19 participating staff clinicians; individual items scored on a six-point scale (0 = Absent, 5 = Very Good); corrected item-scale correlation reflects association of the item with the full scale score when the item is excluded from the full scale score; internal consistency of the six-item HRQ-CM was Cronbach $\alpha = .74$.