Routledge
Taylor & Francis Group

The development of instruments to measure motivational interviewing skill acquisition for school-based personnel

Jason W. Small[a]*, Jon Lee[b], Andy J. Frey[c], John R. Seeley[a] and Hill M. Walker[d]

[a]*Oregon Research Institute, Eugene, OR, USA;* [b]*School of Education, University of Cincinnati, Cincinnati, OH, USA;* [c]*Kent School of Social Work, University of Louisville, Louisville, KY, USA;* [d]*College of Education, University of Oregon, Eugene, OR, USA*

As specialized instructional support personnel begin learning and using motivational interviewing (MI) techniques in school-based settings, there is growing need for context-specific measures to assess initial MI skill development. In this article, we describe the iterative development and preliminary evaluation of two measures of MI skill adapted from the substance abuse field for use in school-based settings. We developed the Video Assessment of Simulated Encounters for School-Based Applications and the Written Assessment of Simulated Encounters for School-Based Applications to evaluate the initial MI skill development of school-based personnel participating in a multi-component, MI training program. Preliminary psychometric evidence supports continued development and refinement of the measures.

Keywords: instrumentation; motivational interviewing competency; training and supervision; treatment fidelity

Even the most skilled practitioners in the fields of mental health, substance abuse, and health care cannot learn motivational interviewing (MI) 'in a few minutes over a pizza' (Rollnick, Miller, & Butler, 2008, p. 177). Development of MI proficiency is a multi-stage process involving development of relational skills, such as empathy and an MI spirit, and technical, MI-consistent methods (Hartzler, Beadnell, Rosengren, Dunn, & Baer, 2010; Miller & Rose, 2009). The path to MI proficiency begins with a willingness to assume a client-centered approach and develop client-centered counseling skills; requires attunement to the language of change, particularly the ability to recognize, reinforce, and elicit language supporting a person's expressed desire to change his or her behavior; and necessitates a sensitivity to when and how to initiate change talk, and recognize and support a person's commitment to that change. MI proficiency culminates in the ability to adeptly and seamlessly integrate MI skills with other evidence-based counseling methods into the client–therapist relationship (Miller & Moyers, 2006).

In this article, we discuss the theoretical distinction between skill acquisition and proficiency and distinguish between training methods that impart initial skill and training methods that facilitate long-term mastery. We then discuss the importance of context-specific MI-trainings and the need for context-specific MI skill and proficiency measurement. Finally, we discuss the initial, iterative development and testing of two measures of MI skill adapted from the substance abuse field for use with school-based mental health professionals (e.g., school social workers, school psychologists, school counselors, and behavior specialists).

---

*Corresponding author. Email: jasons@ori.org

From initial skill acquisition to proficiency

Development of professional competence is incremental and ongoing (Kaslow et al., 2007; Leigh et al., 2007). Approaches to developing and sustaining MI proficiency range from group-based skill acquisition via formal didactic trainings and simulated exercises to informal, individualized, skill refinement in authentic settings. Evidence suggests workshops and multi-day trainings support initial MI skill acquisition as well as MI self-efficacy, interest, knowledge, and intent to use (Madson, Loignon, & Lane, 2009; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004; Walters, Matson, Baer, & Ziedonis, 2005). However, the development and refinement of the complex clinical skills comprising an MI-approach must be maintained through on-going, individualized feedback and coaching (Martino et al., 2011; Miller et al., 2004) and mastered through regular implementation and continuous refinement in day-to-day practice (Rollnick et al., 2008). As depicted in Figure 1, these training processes represent a continuum from didactic training – the transfer of information through written materials and workshops – to competence training – the implementation of an evidence-based practice with fidelity (McHugh & Barlow, 2010). Formal training and workshops offer a vital, foundational introduction to the spirit and skills of MI, especially for participants with limited previous exposure to person-centered approaches (Carpenter et al., 2012; Söderlund, Madson, Rubak, & Nilsen, 2011). Individualized feedback and coaching then reinforce foundational skills, ensure maintenance of implementation fidelity, and minimize potential skill drift or decay (Bennett, Moore, et al., 2007). Finally, continuous, reflective use of MI in an authentic setting facilitates iterative informal learning and skill improvement that promotes sustained and proficient use of MI techniques.

Miller and Moyers (2006) described eight sequential steps to developing MI proficiency that Madson et al. (2009) and Söderlund et al. (2011) grouped into two training phases based on systematic reviews of MI training programs. During phase one trainings, the principles of MI are explicitly taught, discussed, and practiced. These MI principles include developing an MI spirit (stage 1), learning client-centered counseling micro-skills (stage 2), recognizing and reinforcing change talk (stage 3), and developing strategies to roll with client resistance (stage 5). Phase two trainings develop skills needed to elicit and strengthen change talk (stage 4), develop a change plan (stage 6), consolidate commitment (stage 7), and switch flexibly between MI and other evidence-based counseling approaches (stage 8). Whereas phase one trainings focus on skills needed to reduce ambivalence and
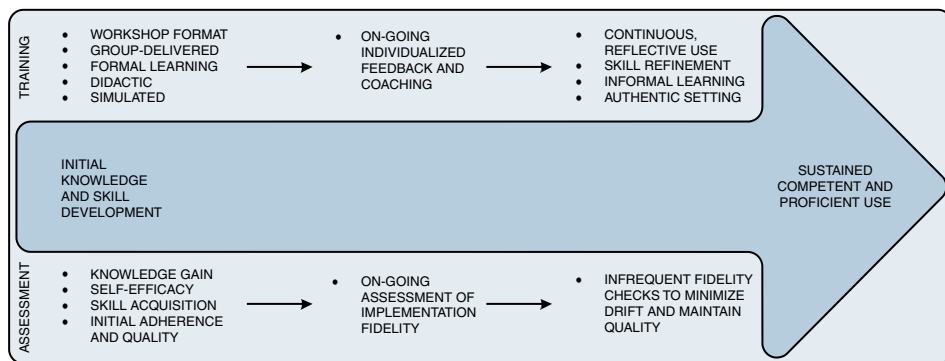

Figure 1. MI training and assessment continuum.

build motivation, phase two trainings develop the skills needed to strengthen and encourage a shift toward behavior change (Söderlund et al., 2011).

MI training for school-based personnel

Literatures on MI training processes in clinical, medical, and community health settings are growing rapidly. Literature reviews of MI trainings (Barwick, Bennett, Johnson, McGowan, & Moore, 2012; Madson et al., 2009; Söderlund et al., 2011) identified more than 35 articles published between 1999 and 2009 reporting varied training models implemented primarily with general and mental health practitioners and clients. Additional MI studies continue to emerge (Bell & Cole, 2008; Cucciare et al., 2012; Martino et al., 2011; Young & Hagedorn, 2012). Despite the wealth of evidence and the variety of training and support models proposed, these reviews have identified only one study examining MI training for school-based personnel (Burke, Da Silva, Vaughan, & Knight, 2005). Burke and her colleagues delivered a single-session training on the principles of MI as part of a project to enhance the screening, assessment, intervention, and referral skills of seven school-based personnel working in secondary schools counseling students with or at-risk of developing substance abuse problems.

Translating MI trainings to school-based settings may be challenging. Person-centered approaches predictive of MI skill development and proficiency (Carpenter et al., 2012; Miller et al., 2004; Söderlund et al., 2011) may be more difficult to develop and sustain for school-based practitioners who may not possess the foundational skills needed to learn the approach. Developing proficiency may prove even more difficult given the limited time school-based practitioners have for professional development and the reduced opportunities they have to apply and refine MI skills during day-to-day interactions (Lee et al., 2014). School-based applications of MI present unique challenges not addressed in health and mental health MI training modules. Unlike applications in settings where a client receives counseling to encourage change in his or her behavior, school-based practitioners utilize an MI approach with teachers and parents to, ultimately, effect change in a student's behavior. Within the consultative relationship, practitioners may use MI techniques to encourage change in teaching or parenting practices; however, the discussion must be framed within the context of how adjustments in practice can mediate change in student behavior. As well, practitioners may use MI techniques within the context of school-based coaching models to increase implementation fidelity of an evidence-based practice, an application unique to school-based practice (Frey et al., 2013). In turn, the need for context-specific training experiences that approximate authentic encounters is paramount. Training programs offering activities and scenarios drawn from contexts with which participants are familiar (e.g., providing school personnel with training and MI materials specific to school-based settings) and providing on-going support in authentic settings facilitates skill development, encourages proficiency, and improves implementation fidelity (Herman, Reinke, Frey, & Shepard, 2014; Miller et al., 2004; Rollnick, Kinnersley, & Butler, 2002).

Context-specific MI skill and proficiency measurement

Researchers and trainers frequently collect self-reported measures of MI efficacy and knowledge, and brief, group-administered outcome measures that assess MI skills such as empathy, reflective listening, and eliciting change talk to evaluate MI trainings. They also use detailed coding systems such as the Motivational Interviewing Treatment Integrity

(MITI) code (Moyers, Martin, Manuel, Hendrickson, & Miller, 2005) to score audio- or video-recorded practice samples. Despite the availability and broad use of these varied measures, many have limited psychometric evidence to support them (Madson et al., 2009). As well, utility would be greatly improved if length and complexity of these measures could be reduced and reliability improved, especially in training and supervisory contexts where efficient, accurate, and reliable assessment is vital.

As Bennett, Roberts, Vaughn, Gibbins, and Rouse (2007) argue, if MI competence is context-bound, assessments too should reflect the setting within which, and clients with whom, professionals implementing MI techniques work. Thus, the development of context-specific MI training and support modules necessitates the creation of new, or adaptation of existing, measures that are (1) sensitive to training effects and (2) predictive of future MI proficiency.

Although context-specific measures may be less relevant during the latter stages of learning when sustained, proficient use increases and skills are generalized across settings and clients, measures of skill development have the potential to provide valuable formative feedback during early training stages and may help identify specific skill deficiencies for targeted support and reinforcement.

Instruments suitable or adapted for school-based settings are not available and, to date, the only systematic attempt to measure the MI-related skills of school-based personnel has been by Frey et al. (2013). Later, we describe the iterative adaptation process and preliminary evaluation of two school-based measures, the Video Assessment of Simulated Encounters for School-Based Applications (VASE-SBA; Lee, Frey, & Small, 2013) and the Written Assessment of Simulated Encounters for School-Based Applications (WASE-SBA; Lee, Small, & Frey, 2013). We developed these instruments to evaluate initial MI skill development of school-based personnel participating in the Motivational Interviewing Training and Support (MITS) module, a multi-component MI training program tailored to personnel working in school settings.

Iterative development and testing of the VASE-SBA and WASE-SBA

We developed the VASE-SBA and WASE-SBA using McCoach, Gable, and Madura (2013) and DeVellis' (2011) recommended steps for scale development. These steps include (1) conceptual definition and literature review, (2) pre-test, (3) expert panel review, and (4) pilot test. Later, we describe the stages of completion of the first three steps in this recommended process and our plan to pilot test two school-based MI skill measures.

Conceptual definition and literature review

In fall 2012, we conducted a review of measures listed on the assessment and coding resource page of the Motivational Interviewing Network of Trainers (MINT) website to identify instruments we could adapt for application in school settings. We conducted the review to identify instruments that (1) assessed MI skill and competence, (2) had been administered previously in early stage MI training and workshop settings, (3) had evidence of short-term change sensitivity, and (4) had at least preliminary evidence to support reliability and validity. We excluded instruments that were overly complex, inefficient, or costly. During our review, we identified eight instruments. We eliminated four instruments due to their complexity, one instrument because it measured only knowledge change, and one instrument designed for sequential analyses. We identified two as promising for adaptation in the context of school-based intervention practice and research: the Helpful

Response Questionnaire (HRQ; Miller, Hedrick, & Orlofsky, 1991) and the Video Assessment of Simulated Encounters-Revised (VASE-R; Rosengren, Baer, Hartzler, Dunn, & Wells, 2005).

The HRQ is a brief (i.e., 15–20 min), free-response measure developed for administration in group settings to measure accurate empathy, the ability to sensitively and accurately infer someone's thoughts, feelings, and struggles (Rogers, 1951). The HRQ consists of six hypothetical client statements. The six written statements simulate communication from individuals ranging from a 15-year-old girl struggling with issues related to peer pressure to a 59-year-old unemployed teacher struggling with his role as a father and his ability to find employment. To assess the ability to generate empathic responses, respondents are asked to write down how they would respond verbally to each simulated communication (e.g., 'Write here what you would say next.'). Raters code each response on a five-point scale ranging from *1* for responses containing no reflection and including a roadblock (Gordon, 2008) to *5* for responses including paraphrasing or inferred meaning and inferred emotion appropriate to the written statement. Responses containing roadblocks, and thereby indicating low levels of accurate empathy, are scored with an item rating below *3* and are used to inform training feedback. Cronbach's α, a measure of the internal consistency of the measure, ranged from 0.92 for pre-training administration to 0.89 for post-training administration. Item-level reliability coefficients ranged from 0.71 to 0.91 and inter-rater reliability for the HRQ total score was 0.93 (Miller et al., 1991).

The VASE-R assesses five MI skills: reflective listening, response to resistance, summarizing, eliciting change talk, and developing discrepancy. The VASE-R consists of three video vignettes in which actors portray substance abusers. Each vignette includes a number of statements by a hypothetical client. After each statement, the instrument administrator pauses the recording and prompts respondents to write a response consistent with a specified MI principle. In total, there are 18 items ($n = 6$ items per vignette). For each vignette, there are five open-ended response items and one multiple-choice item, for which the respondent must also provide a written rationale for the choice made. The VASE-R is appropriate for group delivery and takes approximately 35 min to administer. In addition to the total scale score, there are five subscales: a 4-item *reflective listening* subscale assessing the respondent's ability to generate accurate reflections; a 5-item *responding to resistance* subscale assessing production of non-confrontational responses; a 3-item *summarizing* subscale assessing the generation of summary statements including client ambivalence and change talk; a 3-item *eliciting change talk* subscale assessing skillful production of responses likely to elicit client change talk; and a 3-item *developing discrepancy* subscale assessing identification of utterances likely to enhance motivation for change. Raters code each item on a three-point scale. Raters score an item *0* if the written response is confrontational or engenders resistance, *1* if it is neutral or inaccurately represents the content of the hypothetical client's statement, or *2* if the response reflects appropriate use of the MI skill being assessed. Rosengren, Hartzler, Baer, Wells, and Dunn (2008) report a Cronbach's α of 0.85 for the VASE-R total score with internal consistency for the subscales ranging from 0.44 for *developing discrepancy* to .73 for *summaries*. Item-level inter-rater reliability coefficients by subscale were in the acceptable range across two reported studies with intra-class correlations (ICCs) ranging from 0.41 to 0.96. As evidence of concurrent validity, Rosengren et al. (2008) report that the HRQ total score was correlated with VASE-R total scores ($r = 0.50$) and subscale scores with correlations ranging from 0.23 for the *developing discrepancy* subscale to 0.45 for the *reflective listening* subscale.

Instrument adaptation and development

We modified the HRQ and VASE-R written samples and video vignettes to reflect scenarios and concerns that school-based mental health professionals would likely encounter.

Initial adaptations

The HRQ presents six hypothetical, written client statements to which respondents are asked to provide a one or two sentence reflective response. To begin, we iteratively adapted these written prompts. Although we attempted in early iterations to attend as closely as possible to the original HRQ statements, the versions we developed for pre-testing focused on (1) providing a more detailed and varied description of the hypothetical speaker and (2) providing a written statement that maintained the spirit of the original but reflected a concern that might realistically be raised in school-based settings. For the VASE-R, we adapted client background statements and video vignettes using a similar iterative editing and revision process to align the measure with scenarios relevant to a school-based context. For pre-testing, we modified the original structural organization of the VASE-R only slightly, with the addition of one prompt per vignette (a request for an affirmation of the vignette actor). At pre-testing the adapted structure included three vignettes with seven response items per vignette, and two introductory sample items ($n = 23$ total items). We also consolidated the VASE-R's general scoring rules, provided specific guidelines and examples to facilitate interpretation of these rules, and created a scoring sheet to streamline scoring.

Pre-testing

We tested the adapted HRQ and VASE-R as pre–post outcomes for a pilot study of MI training modules for school-based personnel. We then further adapted the measures and their respective scoring protocol using a two-step iterative process. During step one, two coders evaluated participants' open-ended responses at baseline and compiled detailed notes to inform subsequent revisions. Then, at step-two, we further refined and clarified scoring procedures based on coding of post-training data.

For the HRQ, coders used the original scoring procedures developed by Miller et al. (1991). After iterative testing and coding, we made minor amendments to the coding protocol to facilitate reliable coding of open-ended responses. Specifically, we explicitly linked the description of scoring anchors to the presence (or absence) of specific reflective practices. For example, we linked a rating of 3 to simple reflections, a rating of 4 to complex reflections, and a rating of 5 to statements that infer potential parent or teacher behavior change (e.g., linking of individual values, labeling specific behavior changes, or inference to preparatory or mobilizing change talk).

Following pre-testing, we revised the VASE-R more extensively. We aligned the subscales with Miller and Rollnick's (2012) refined MI3 conceptualization, which identifies the MI micro skills as open-ended questions, affirmation, reflections, and summaries (OARS). In response to feedback from participants who raised concerns about administration time, we reduced the VASE-R to two vignettes with 12 response items per vignette. Table 1 contains the background descriptions for the two retained vignettes adapted for school-based administration. Although the reduction in vignettes actually increased the number of items (from 23 to 24), the alteration reduced administration time because the overall number of prompts spoken by each vignette actor was reduced from 23

Table 1. VASE-SBA client background descriptions.

Client	Client background description
Lisa	Lisa is a forty-three-year-old female teacher of preschoolers who is midway through her twentieth year of teaching at your school. She has invited you to work with her, as she feels she has 'lost her touch' when it comes to classroom management. She believes that families have become more lenient with their children over the years, and feels this has resulted in the increase in behavior problems the school has seen since she first began teaching. Lisa relies on the traditional 'color card' system in her classroom, which uses a green, yellow, and red card that the children 'flip' after breaking the rules.
Bailey	Bailey is a thirty-two-year-old father of three boys. Each of his children has attended your elementary school. Now, Bailey's youngest child, Elise who is in pre-k, is having behavior problems in school. As the school social worker, you've invited Bailey in to talk about the behavior problems that have been reported by the teacher.

to 15. We retained the VASE-R's original 3-point scale scoring format; however, we altered the wording of item anchors to align with MI3 language (e.g., resistance to discord): ratings of 1 or 2 (containing roadblocks) may elicit/reinforce sustain talk or engender discord, while a rating of 3 (simple reflection) is neutral, and ratings for 4 or 5 (complex reflections and identifying behavior change) motivational.

Expert review

In August 2013, we consulted via e-mail with the senior authors of each instrument to obtain feedback on the adapted measures. We sent each author a detailed record of changes made and our rationale for doing so, requested suggestions for further improvement, and initiated discussion about the appropriate naming and citation of each adapted measures. These authors provided detailed suggestions to further improve our adaptations and align them with more recent MI3 conceptualizations.

Based on expert feedback on the HRQ adaptations, we further revised two of six statements (items 2 and 5), so the measure included positive statements rather than just negative statements. The current version, as reported in Table 2, now includes four negative and two positive statements reported by four hypothetical teachers of varying age, gender, and teaching experience, and two parents, a mother and a father.

Based on expert feedback on the VASE-R adaptations, we adjusted prompts to provide increased use of reflections and open-ended questioning early in each vignette and to increase opportunities for the use of reflections, affirmations, and summaries near the middle and end of each vignette. Finally, we further refined and clarified scoring anchors.

Method

Participants and procedures

Twelve consultants working in early childhood programs serving children four-years-old and younger in Louisville, Kentucky participated in a pilot study of an MI training module for school-based personnel. The 12 participants reported job titles of curriculum resource teacher (25%), disability liaison (25%), special education resource teacher (25%), and social worker (25%). On average, participants had worked 9 years (SD = 10.6) in their current positions and had been teaching 15 years (SD = 9.4). Half of the participants reported holding a Master's degree in education, counseling, or social work. The majority

Table 2. WASE-SBA hypothetical client statements.

Item	Written statement
1	*A 29-year-old female teacher, who is in her seventh year of teaching at your school, says*: 'Yesterday Bobby came to school, obviously agitated. When I asked him what was wrong, he would not look at me or respond. As the morning progressed his agitation got worse. When I asked him to participate in our group meeting, he yelled out and began to shove the other children. It got worse when I intervened, and he tried to hit me! I just don't know what to do!'
2	*A 55-year-old male teacher, who has worked for the school for 25 years says*: 'Justine's mother can really make a person mad. She's always coming to school and bothering us or complaining about things that really never happened. It can take time away from the kids, but I've found that giving her a classroom responsibility – like recording homework or running a small discussion group – helps her stay appropriately engaged!'
3	*A 23-year-old female teacher, who is in her first year of teaching, says*: 'I'm really mixed up. A lot of the teachers here do things I don't agree with and the Principal tells us not to do. Like, they tell me to send Fredrick to the timeout room when he misbehaves, but we're supposed to handle misbehavior in our rooms. I want to fit in, but I don't know what would happen if I went along with them.'
4	*A 35-year-old mother says*: 'My Maria is a good girl. She's never been in trouble, but I worry about her. Especially, when she is with her father (we are divorced) she does things that she knows I am not in favor of. She just had her ears pierced without asking me! I feel like her father is too easy on her, and now I'm "mean mommy!"'
5	*A 19-year old father says*: 'I really feel awful. Yesterday I yelled at Andy again. I just lose my temper and forget all about the right ways to discipline him. My own parents know I am struggling with him, and they sat me down to talk about it – it was very uncomfortable – but they've helped me see that I've got to change this, but I don't know how.'
6	*A 49-year-old teacher, who has taught at your school for 27 years, says*: 'My work as a teacher just doesn't seem worth it any more. I'm lousy with discipline. The kids hate me, because I am so mean. But these kids are just rotten! Sometimes I wonder whether they've got parents at home at all!'

were female (91%) with a mean reported age of 48 years (SD = 9.0). Participants reported their ethnicity as African American (25%) and Caucasian (75%). One participant reported taking a masters-level MI course through a school of social work program. The remaining participants had no previous MI training or experience.

The University of Louisville Institutional Review Board approved all procedures. The investigators recruited the 12 consultants who participated in the study. MI training sessions were held at a school district training facility. Recruited participants provided informed consent, agreed to attend all training sessions, and agreed to complete baseline and post-training assessment packets. Participants received three $50 gift cards: one for completing baseline assessment measures, one for participating in training, and one for completing post-training assessment measures. The early childhood program employing participating consultants did not receive compensation for their employees' participation; however, the training sessions, coach support, and training materials were provided as part of a mental health services contract administered by the third author.

After obtaining informed consent, participants completed the adapted versions of the HRQ and VASE-R as part of a larger baseline assessment battery. All 12 consultants then participated in the MITS module, a multi-component MI training program tailored to personnel working in school settings. The MITS module uses didactic and interactive

teaching methods such as lectures, discussion of key concepts, modeling (through video and live demonstration), and role-playing. The core content of the module focuses on Miller and Rollnick's (2012) MI processes such as MI micro skills (i.e., OARS), evoking change talk, increasing or decreasing sustain talk, and planning for change. The 19-hour training module is delivered over a three-month period and includes (1) five 3-hour workshops, (2) a 1-hour, unstructured group discussion, and (3) three, 1-hour individualized coaching sessions during which participants receive feedback on their use of MI with a parent or teacher. After completing the training module, participants again completed the adapted HRQ and VASE-R as part of a larger post-training assessment battery.

Two coders, Drs Frey and Lee, scored the HRQ and VASE-R data. Reliability scoring procedures varied by measures. Given the complexity of the VASE-R scoring procedures and their application to new scenarios, the coders worked in tandem to score the responses. They took detailed notes and discussed coding discrepancies, interpretive differences, and other issues related to applying a pre-existing coding and scoring procedure to newly adapted, open-ended response categories specific to a new setting and context. Although we acknowledge that this approach compromises inter-rater reliability and artificially inflates the agreement statistics reported later for the VASE-R, we believe this was necessary given the current stage of measurement development.

Statistical analysis

We examined two measures of reliability: internal consistency, an index of consistency across items, and inter-rater reliability, an index of consistency across raters. Specifically, we examined internal consistency for each rater using Cronbach's coefficient alpha (α) and inter-rater reliability using the ICC. We utilized Cicchetti's (1994) recommendations to assess the sufficiency of ICCs. To assess each measure's sensitivity to change, we examined within-subject effects in an analysis of variance framework using the general linear model procedure in SPSS 19.

Results

Reliability

For the HRQ, coefficient α was 0.71 and 0.76, respectively, for the two raters. For the VASE-R scale, coefficient α was 0.81 and 0.77 for the two raters. For HRQ item-level, ICCs were all in the acceptable range (i.e., ICC > 0.40). Inter-rater reliability was lowest for items 1 and 2 (ICCs = 0.58 and 0.54, respectively) with considerably higher ICCs for the remaining four items (mean ICC = 0.90; range = 0.82–0.95). For the HRQ total score, inter-rater reliability was excellent (ICC = 0.92). ICCs for the VASE-R subscales ranged from 0.79 for the change talk subscale to 0.99 for the reflective listening and developing discrepancy subscales. For the VASE-R total score, the ICC was 0.99.

Item and scale statistics

Table 3 summarizes item-level means for the six HRQ items for the two raters. Baseline item statistics are similar for the two raters. For items 1, 2, and 4, coded ratings for both informants ranged from a minimum of 1 to a maximum of 2. Scores for item 5 ranged from 1 to 3 and from 1 to 4 for item 6 for both raters. Rating variability was similar

Table 3. Baseline and post-training HRQ item-level statistics by rater.

	Baseline		Post-training	
Item	Rater 1 M (SD)	Rater 2 M (SD)	Rater 1 M (SD)	Rater 2 M (SD)
1	1.25 (0.45)	1.33 (0.49)	2.50 (1.24)	2.58 (1.16)
2	1.33 (0.49)	1.08 (0.29)	3.25 (1.22)	3.25 (1.22)
3	1.58 (0.90)	1.42 (0.51)	3.25 (1.14)	3.25 (1.14)
4	1.42 (0.51)	1.50 (0.52)	3.33 (0.98)	3.33 (0.98)
5	1.42 (0.67)	1.33 (0.65)	3.00 (1.48)	3.00 (1.48)
6	2.00 (1.28)	1.83 (1.19)	3.27 (1.27)	3.27 (1.27)

across informants for the aforementioned items; though, for item 3, rater 1's scores were more variable ranging from 1 to 4 on the 5-point scale as compared with rater 2's scores, which ranged from 1 to 2. Mean item-level scores fall into the middle of the 5-point scoring range for five of six items. Only item 1 has a mean rating below 3. There was a high level of agreement in post-training scores across the two raters with uniform, item-level agreement for five of six items. Based on rater 1's data, HRQ total scores increased from 9.0 (SD = 3.0) to 18.3 (SD = 3.2). The within-subject partial r effect size was 0.92.

Baseline VASE-R total scores and subscale scores reported in Table 4 were comparable across raters. VASE-R total scores can range from 0 to 42. For rater 1, total scores ranged from 7 to 27 and from 6 to 27 for rater 2. Subscale scores were also comparable across raters with mean subscale scores falling below the midpoint of possible scores. VASE-R post-training scores were also comparable across raters. Post-training total scores ranged from 14 to 29 and 12 to 28 for raters 1 and 2, respectively. Mean total scores were above the midpoint of possible scores indicating improvement in MI skills at post training. This pattern held for the reflective listening, responding to resistance, summaries, and eliciting change talk subscales but not for the developing discrepancy and affirmations subscales. Thus, preliminary data suggest the VASE-R scale and four of six subscales are sensitive to training effects. The within-subject partial r effect size for the VASE-R total score was 0.90. There were large within-subject effects for the reflective listening ($r_{part} = 0.88$), responding to resistance ($r_{part} = 0.80$), and summaries subscales ($r_{part} = 0.80$), a moderate effect for the eliciting change talk ($r_{part} = 0.52$), and little to no effect on developing discrepancy ($r_{part} = 0.07$) and affirmations ($r_{part} = 0.07$).

Table 4. Baseline and post-training VASE-R descriptive statistics by rater.

	Baseline		Post-training	
Subscale	Rater 1 M (SD)	Rater 2 M (SD)	Rater 1 M (SD)	Rater 2 M (SD)
Total score	14.6 (6.6)	14.4 (7.2)	23.1 (5.0)	22.3 (4.8)
Reflective listening	2.0 (2.1)	1.9 (2.0)	5.0 (1.4)	4.9 (1.4)
Responding to resistance	2.7 (2.5)	2.9 (2.5)	5.4 (1.8)	4.9 (1.9)
Summaries	1.5 (1.7)	1.5 (1.6)	3.3 (1.2)	3.4 (1.3)
Eliciting change talk	2.0 (1.2)	1.8 (1.6)	2.8 (1.6)	2.7 (1.4)
Developing discrepancy	3.4 (1.5)	3.3 (1.4)	3.5 (1.2)	3.3 (1.3)
Affirmations	3.0 (1.7)	3.0 (1.5)	3.1 (1.4)	3.1 (1.4)

Table 5. Correlations among HRQ total and item scores and VASE-R scale and subscale scores.

VASE-R	HRQ						
	Total	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Total score	0.89***	0.46	0.66*	0.58	0.56	0.42	0.79**
Reflective listening	0.63*	0.27	0.68*	0.56	0.38	0.29	0.63*
Responding to resistance	0.86**	0.48	0.62*	0.65*	0.62*	0.41	0.78**
Summaries	0.46	0.12	0.13	0.19	− 0.02	0.19	0.33
Eliciting change talk	0.42	0.25	0.22	0.52	0.75**	0.13	0.29
Developing discrepancy	0.44	0.21	0.17	− 0.27	− 0.06	0.26	0.30
Affirmations	0.31	0.30	0.42	0.41	0.41	0.18	0.39

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

Concurrent validity

As preliminary evidence of concurrent validity and to better understand the relationship between the HRQ and VASE-R as adapted measures of MI skill, we examined the relationship between the HRQ total and item scores and VASE-R scale and subscale scores. Table 5 provides a correlation matrix of the two measures. VASE-R total scores and HRQ total scores were highly correlated ($r = 0.89$). The VASE-R was moderately correlated with HRQ items 1–5 and highly correlated with item 6. Correlations among HRQ total scores and VASE-R subscale scores were highest for responding to resistance and reflective listening. HRQ total scores were modestly correlated with the remaining VASE-R subscales.

Discussion

Increasingly, school-based personnel are using MI to encourage teacher, parent, and adolescent behavior change (Dishion, Stormshak, & Siler, 2010; Frey et al., 2014; Reinke, Lewis-Palmer, & Merrell, 2008). As the use of MI in school-based settings increases, so too does the need for context-specific MI training and assessment systems. Unfortunately, school-based MI research has been limited because efficient, sensitive, reliable, and valid measures of MI skill and competency for school-based personnel do not exist. This article contributes to the knowledge base by providing an overview of the MI training and competency literature and describing initial steps in an iterative process to develop two, context-specific measures of MI skill acquisition for school-based professionals. We believe the review and integration of a literature base representing diverse practice settings and a range of professions provides school-based researchers and practitioners with valuable information to inform the effective transfer of MI to educational settings. We have discussed the theoretical distinction between skill acquisition and proficiency, and the differences between training methods that impart initial skill and those that facilitate long-term mastery. At a minimum, this synthesis suggests that fairly intensive, context-specific training is required for school-based mental health providers to learn initial MI skills and apply these skills proficiently in practice settings.

The primary contribution of this study is the detailing of an initial, iterative process used to develop two measures of MI skill. Specifically, we described our process for completing McCoach et al. (2013) and DeVellis' (2011) recommended steps for scale development: (1) conceptual definition and literature review, (2) pre-test, and (3) expert panel review. We identified the HRQ (Miller et al., 1991) and the VASE-R (Rosengren

et al., 2005) as promising for adaptation to school-based research and practice settings, iteratively adapted the measures, and modified the names and citations of the measures with permission from the original authors. We believe the VASE-SBA (Lee, Frey, & Small, 2013) and WASE-SBA (Lee, Small, & Frey, 2013) are the first context-specific MI skill measures developed for school-based settings. Although these measures are still under development, our results provide preliminary evidence the psychometric properties for the measures will be acceptable.

The development of professional competence, be it in a technique such as MI or a broader field of study, is incremental and ongoing (Kaslow et al., 2007). Competence is a complex, multi-dimensional construct involving the integration of knowledge, skills, and attitudes (Leigh et al., 2007; Lichtenberg et al., 2007). In turn, it may be difficult to assess within the context of contrived training environments. Although a measure such as the MITI offers a broad assessment of MI proficiency via global ratings and technical skill tallies, coding data using the MITI is costly, time consuming, and requires collection of audio-recorded samples. In contrast, the WASE-SBA and VASE-SBA are brief, efficient, and cost-effective measures that can be collected prior to and following trainings via either a group-delivered format or web-based interface. We do not propose that the WASE-SBA and VASE-SBA should replace more extensive coding systems. Instead, we believe the WASE-SBA and VASE-SBA should be conceptualized within a broader assessment continuum as depicted in Figure 1. In this context, these instruments offer promise as measures of short-term change in MI skill acquisition and, pending further evaluation, may prove predictive of future MI proficiency.

The WASE-SBA and VASE-SBA are part of a broader MI training and assessment system we are currently developing and evaluating. The assessment system includes content-specific measures of MI knowledge gain, self-reported measures of perceived proficiency, measures of training motivation and MI self-efficacy, and a version of the MITI adapted for school-based settings (Frey et al., 2014). The WASE-SBA, VASE-SBA, and MI knowledge measures capture short-term knowledge and skill development. We then use MITI data to inform and guide individualized coaching and feedback sessions and document the participant's incremental movement toward MI proficiency and competency. Coaches use the participant-reported measures of perceived proficiency, motivation, and self-efficacy to further inform feedback sessions. The trainer uses these data to (1) identify discrepancies between perceived and measured (i.e., data from the MITI) ability, (2) highlight points of agreement (i.e., triangulation of coach, participant, and MITI data on key development and skill areas), (3) encourage self-reflection during future MI practice, and (4) discuss fluctuations in motivation and self-efficacy that might impede continued skill gain.

As recommended by the APA's Task Force on Assessment of Competence in Professional Psychology (Kaslow et al., 2007), we have conceptualized the training and assessment system to facilitate on-going incremental skill gain, provide formative and summative feedback, and to encourage the self-reflection necessary for continued growth after completion of MI training. We believe this model offers a cost-effective approach to training and evaluation as it prioritizes the use of efficient, low-cost measures such as the WASE-SBA and VASE-SBA to examine initial skill gain and more detailed assessment tools to capture each participant's movement toward MI competence. Although currently used for in-service training, the WASE-SBA and VASE-SBA could also be used to assess skill acquisition for pre-service training, to promote self-directed, ongoing reflection and skill maintenance of trained practitioners, or in the context of professional learning communities (PLCs). For example, the WASE-SBA and VASE-SBA could be

administered and scored within a PLC to facilitate discussion and feedback among participants on specific MI relational or technical skills.

These measures are an important step toward establishing the effectiveness of MI-based professional development activities, and examining how and for whom MI trainings and the use of MI techniques are effective. From a theoretical standpoint, practitioners use MI to increase change talk, decrease sustain talk, and, ultimately, encourage behavior change. Thus, researchers could use these measures to examine whether change in practitioners' MI relational skills mediates increases in client change talk or decreases in client sustain talk. We believe school-based researchers also could use these measures to examine whether, for example, baseline empathy skills moderate MI proficiency (Carpenter et al., 2012; Miller et al., 2004; Söderlund et al., 2011).

Although the psychometric analysis presented in this article does not provide sufficient evidence to argue the WASE-SBA and VASE-SBA are valid and reliable measures, the encouraging results support further development and refinement. Currently, we are pursuing funding for a large-scale validation study with a representative sample of school mental health professionals. We will use data from this study to further (1) examine reliability evidence, including inter-rater reliability and internal consistency; (2) develop additional evidence to support the validity of the measures including examination of the factor structure of each measure using confirmatory factor analysis and item response theory; and (3) establish a normative database and competency benchmarks. In addition, we are planning to administer these measures to 100 early childhood home visitation professionals as part of an MI training effort currently underway.

The primary limitations of this study are the small sample size, lack of independence in the coding procedures, and lack of a 'gold standard' measure for reporting preliminary evidence of concurrent validity. Study findings are based on data from a small pilot study of the MITS training program. Findings would be greatly improved with a much larger sample that would ensure stability and accuracy of reported test statistics. The iterative development process precluded independent coding of data which artificially inflated reported agreement statistics.

We acknowledge that inter-rater reliability statistics must be interpreted with caution and will require detailed examination in future iterations. Adapting measures to new contexts can be complicated in the absence of existing 'gold standard' measures to which newly developed measures can be compared. Although the scope of this study limited opportunities to collect additional measures, preliminary evidence of concurrent validity would be improved by collection of a more detailed coding system such as the MITI.

In conclusion, this study contributes to the knowledge base by distinguishing between MI skill and MI competency measures, expressing the need for context-specific measurement, and describing an iterative development process for two context-specific MI measures for use with school mental health professionals. Future MI research in school-based settings will greatly benefit from the continued development and adaptation of context-specific MI skill and competency measures to evaluate the impact of trainings, the skill development of personnel, and to monitor proficiency and skill maintenance over time.

References

Barwick, M. A., Bennett, L. M., Johnson, S. N., McGowan, J., & Moore, J. E. (2012). Training health and mental health professionals in motivational interviewing: A systematic review. *Children and Youth Services Review*, *34*, 1786–1795. doi:10.1016/j.childyouth.2012.05.012

Bell, K., & Cole, B. (2008). Improving medical students' success in promoting health behavior change: A curriculum evaluation. *JGIM: Journal of General Internal Medicine*, *23*, 1503–1506. doi:10.1007/s11606-008-0678-x

Bennett, G. A., Moore, J., Vaughan, T., Rouse, L., Gibbins, J. A., Thomas, P., . . . Gower, P. (2007). Strengthening motivational interviewing skills following initial training: A randomised trial of workplace-based reflective practice. *Addictive Behaviors*, *32*, 2963–2975. doi:10.1016/j.addbeh.2007.06.013

Bennett, G. A., Roberts, H. A., Vaughan, T. E., Gibbins, J. A., & Rouse, L. (2007). Evaluating a method of assessing competence in motivational interviewing: A study using simulated patients in the United Kingdom. *Addictive Behaviors*, *32*, 69–79. doi:10.1016/j.addbeh.2006.03.022

Burke, P. J., Da Silva, J. D., Vaughan, B. L., & Knight, J. R. (2005). Training high school counselors on the use of motivational interviewing to screen for substance abuse. *Substance Abuse*, *26*, 31–34. doi:10.1300/J465v26n03_07

Carpenter, K. M., Cheng, W. Y., Smith, J. L., Brooks, A. C., Amrhein, P. C., Wain, R. M., & Nunes, E. V. (2012). 'Old dogs' and new skills: How clinician characteristics relate to motivational interviewing skills before, during, and after training. *Journal of Consulting and Clinical Psychology*, *80*, 560–573. doi:10.1037/a0028362

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cucciare, M. A., Ketroser, N., Wilbourne, P., Midboe, A. M., Cronkite, R., Berg-Smith, S. M., & Chardos, J. (2012). Teaching motivational interviewing to primary care staff in the Veterans Health Administration. *Journal of General Internal Medicine*, *27*, 953–961. doi:10.1007/s11606-012-2016-6

DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

Dishion, T. J., Stormshak, E., & Siler, C. (2010). An ecological approach to interventions with high-risk students in schools: Using the Family Check-Up to motivate parents' positive behavior support. In M. R. Shinn H & M. E. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 101–125). Bethesda, MD: National Association of School Psychologists.

Frey, A. J., Lee, J., Small, J. W., Seeley, J. R., Walker, H. M., & Feil, E. G. (2013). Motivational Interviewing Navigation Guide: A process for enhancing teacher's motivation to adopt and implement school-based interventions. *Advances in School Mental Health Promotion*, *6*, 158–173. doi:10.1080/1754730X.2013.804334

Frey, A. J., Small, J. W., Lee, J., Walker, H. M., Seeley, J. R., Feil, E. G., & Golly, A. (2014). Expanding the range of the First Step to Success intervention: Tertiary-level support for children, teachers, and families. *Early Childhood Research Quarterly*. doi:10.1016/j.ecresq.2014.05.002

Gordon, T. (2008). *Parent effectiveness training: The proven program for raising responsible children*. New York, NY: Three Rivers Press.

Hartzler, B., Beadnell, B., Rosengren, D. B., Dunn, C., & Baer, J. S. (2010). Deconstructing proficiency in motivational interviewing: Mechanics of skilful practitioner delivery during brief simulated encounters. *Behavioural & Cognitive Psychotherapy*, *38*, 611–628. doi:10.1017/S1352465810000329

Herman, K., Reinke, W., Frey, A. J., & Shepard, S. (Eds.). (2014). *Motivational interviewing in schools: Strategies to engage parents, teachers, and students*. New York, NY: Springer Publishing.

Kaslow, N. J., Rubin, N. J., Bebeau, M. J., Leigh, I. W., Lichtenberg, J. W., Nelson, P. D., . . . Smith, I. L. (2007). Guiding principles and recommendations for the assessment of competence. *Professional Psychology: Research and Practice*, *38*(50), 441–451. doi:10.1037/0735-7028.38.5.441

Lee, J., Frey, A. J., & Small, J. W. (2013). *The video assessment of simulated encounters – School-based applications*. Cincinnati, OH: University of Cincinnati.

Lee, J., Frey, A. J., Walker, H., Golly, A., Seeley, J., Small, J., & Feil, E. (2014). Motivational interviewing in support of teacher behavior change. In E. McNamara (Ed.), *Motivational interviewing with children and young people II: Issues and further applications* (pp. 83–102). Southport, Merseyside, UK: Positive Behaviour Management.

Lee, J., Small, J. W., & Frey, A. J. (2013). *Written assessment of simulated encounters – School-based application*. Cincinnati, OH: University of Cincinnati.

Leigh, I. W., Smith, I. L., Bebeau, M. J., Lichtenberg, J. W., Nelson, P. D., Portnoy, S., . . . Kaslow, N. J. (2007). Competency assessment models. *Professional Psychology: Research and Practice*, *38*, 463–473.

Lichtenberg, J. W., Portnoy, S. M., Bebeau, M. J., Leigh, I. W., Nelson, P. D., Rubin, N. J., . . . Kaslow, N. J. (2007). Challenges to the assessment of competence and competencies. *Professional Psychology: Research and Practice*, *38*, 474–478.

Madson, M. B., Loignon, A. C., & Lane, C. (2009). Training in motivational interviewing: A systematic review. *Journal of Subsance Abuse Treatment*, *36*, 101–109. doi:10.1016/j.jsat.2008.05.005

Martino, S., Ball, S. A., Nich, C., Canning-Ball, M., Rounsaville, B. J., & Carroll, K. M. (2011). Teaching community program clinicians motivational interviewing using expert and train-the-trainer strategies. *Addiction*, *106*, 428–441. doi:10.1111/j.1360-0443.2010.03135.x

McCoach, B. D., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain* (3rd ed.). New York, NY: Springer.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist*, *65*, 73–84.

Miller, W. R., Hedrick, K. E., & Orlofsky, D. R. (1991). The Helpful Responses Questionnaire: A procedure for measuring therapeutic empathy. *Journal of Clinical Psychology*, *47*, 444–448. doi:10.1002/1097-4679(199105)47:3<444:AID JCLP2270470320>3.0.CO;2-U

Miller, W., & Moyers, T. (2006). Eight stages in learning motivational interviewing. *Journal of Teaching in the Addictions*, *5*, 3–17. doi:10.1300/J188v05n01_02

Miller, W. R., & Rollnick, S. R. (2012). *Motivational interviewing: Helping people change (applications of motivational interviewing)* (3rd ed.). New York, NY: The Guilford Press.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist*, *64*, 527–537. doi:10.1037/a0016830

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, *72*, 1050–1062.

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, *28*, 19–26.

Reinke, W. M., Lewis-Palmer, T., & Merrell, R. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*, 315–332.

Rogers, C. R. (1951). *Client-centered therapy*. Boston, MA: Houghton-Mifflin.

Rollnick, S., Kinnersley, P., & Butler, C. (2002). Context-bound communication skills training: Development of a new method. *Medical Education*, *36*, 377–383.

Rollnick, S., Miller, W. R., & Butler, C. C. (2008). *Motivational interviewing in health care: Helping patients change behavior*. New York, NY: Guilford Press.

Rosengren, D. B., Baer, J. S., Hartzler, B., Dunn, C. W., & Wells, E. A. (2005). The video assessment of simulated encounters (VASE): Development and validation of a group-administered method for evaluating clinician skills in motivational interviewing. *Drug & Alcohol Dependence*, *79*, 321–330. doi:10.1016/j.drugalcdep.2005.02.004

Rosengren, D. B., Hartzler, B., Baer, J. S., Wells, E. A., & Dunn, C. W. (2008). The video assessment of simulated encounters-revised (VASE-R): Reliability and validity of a revised measure of motivational interviewing skills. *Drug & Alcohol Dependence*, *97*, 130–138. doi:10.1016/j.drugalcdep.2008.03.018

Söderlund, L. L., Madson, M. B., Rubak, S., & Nilsen, P. (2011). A systematic review of motivational interviewing training for general health care practitioners. *Patient Education & Counseling*, *84*, 16–26. doi:10.1016/j.pec.2010.06.025

Walters, S. T., Matson, S. A., Baer, J. S., & Ziedonis, D. M. (2005). Effectiveness of workshop training for psychosocial addiction treatments: A systematic review. *Journal of Substance Abuse Treatment*, *29*, 283–293. doi:10.1016/j.jsat.2005.08.006

Young, T. L., & Hagedorn, W. B. (2012). The effect of a brief training in motivational interviewing on trainee skill development. *Counselor Education and Supervision*, *51*, 82–97.