Running Head: FIDELITY OF MOTIVATIONAL INTERVIEWING

Fidelity of Motivational Interviewing in School-based Intervention and Research

Jason W. Small[1], Andy Frey[2], Jon Lee[3], John R. Seeley,[4] Terrance M. Scott,[2] & Margaret H. Sibley[5]

[1]Oregon Research Institute, 1776 Millrace Drive, Eugene, OR 97403-1983, USA. Email- jasons@ori.org; Phone 541-484-2123.

[2]University of Louisville, Kent School of Social Work, Patterson Hall, Louisville, KY 40292, USA. Emails- afrey@louisville.edu, t.scott@louisville.edu; Phone- 502-852-0431.

[3]University of Northern Arizona College of Education. PO Box 5774, Flagstaff, AZ 86001, USA. Email- jon.lee@nau.edu, Phone- 928-523-1098.

[4]University of Oregon, College of Education, Eugene, OR 97403. Email: jseeley@uoregon.edu)

[5] University of Washington/Seattle Children's Hospital/ Center for Child Health, Behavior, and Development, 2001 8th Ave., Suite 400, Seattle, WA 98121, USA. Email- Margaret.Sibley@seattlechildrens.org.

Correspondence concerning this article should be addressed to Jason W. Small, Oregon Research Institute, Eugene, OR 97403; Email: jasons@ori.org.

Funding: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R324A150179 to the University of Louisville. The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education.

Conflict of Interest: The authors declare that they have no conflicts of interest.

This report/article was made possible, in part, by the support of the Jefferson County, Public Schools (Louisville, Kentucky) and Greater Clark County Schools (Jeffersonville, Indiana). Opinions contained in this article reflect those of the authors and do not necessarily reflect those of the school districts.

Published in *Prevention Science* 14 September 2020

Small, J. W., Frey, A., Lee, J., Seeley, J. R., Scott, T. M., & Sibley, M. H. (2020). Fidelity of motivational interviewing in school-based intervention and research. *Prevention Science.* Advance online publication. https://doi.org/10.1007/s11121-020-01167-77

Abstract

Educational researchers and school-based practitioners are increasingly infusing Motivational Interviewing (MI) into new and existing intervention protocols to provide support to students, parents, teachers, and school administrators. To date, however, the majority of the research in this area has focused on feasibility of implementation rather than fidelity of implementation. In this manuscript, we will present MI fidelity data from 245 audio-recorded conversations with 113 unique caregivers and 20 coaches, who implemented a school-based, positive parenting intervention. The aggregate fidelity scores across coaches, parents, and sessions provide evidence the training and support procedures were effective in assisting school-based personnel to implement MI with reasonable levels of fidelity in practice settings. Further, results suggest that MI fidelity varied between sessions and coaches and that within-coach variation (e.g., session-level variation in the quality of MI delivered) greatly exceeded between-coach variation. Implications for practice and future research are discussed.

Keywords: Motivational interviewing; School-based; Intervention fidelity; Treatment integrity, parent support; MITI

Fidelity of Motivational Interviewing in School-based Intervention and Research

The benefits of school-based mental health (SBMH) services are well documented. SBMH service delivery extends access to children and youth who otherwise might not be reached; mitigates stigma associated with mental health needs; encourage service provision in natural environments; supports student learning and academic success; and helps increase and maintain school safety (Hoover & Mayworn, 2017; Macklem, 2014). Increasingly, schools are delivering SBMH services within multitiered systems of support which enable efficient delivery of a continuum of evidence-based supports and services but also require these supports and services be delivered with fidelity (Weist et al., 2018).

Successful delivery of evidence-based mental health treatment practices depend, in part, on fidelity or the extent to which practitioners deliver evidence-based programs and practices as prescribed or intended (Sanetti & Kratochwill, 2009). Fidelity, a key implementation outcome (Lewis et al., 2017), is a multi-dimensional construct targeting – most frequently – adherence to program protocol, quality of delivery, and dosage but also extending within broader conceptualizations to program differentiation and participant involvement (Proctor et al., 2011). Collection and examination of fidelity data is central to implementation efforts because it provides evidence of proficient delivery and helps prevent drift across time. Numerous barriers such as fewer resources, lack of time, and limited support complicate fidelity monitoring in school-based settings (Macklem, 2014). Yet, even for rigorous research studies in which resource constraints are less of an impediment, the collection and reporting of fidelity data is inconsistent.

Monitoring fidelity of motivational interviewing (MI) is particularly challenging given most MI fidelity monitoring systems involve detailed coding of verbal interactions. MI is an

evidence-based, collaborative communication style used to explore an individual's motivation for, and commitment to, specific and targeted behavior change (Miller & Rollnick, 2012). Increasingly, educational researchers and school-based practitioners are infusing MI into new and existing intervention protocols. Researchers are examining the use of MI with students, with families, within school-based problem-solving teams, and to improve teachers' implementation of evidence-based practices. To date, however, efforts to examine the implementation of MI in school-based settings have focused primarily on *feasibility* of implementation rather than *fidelity* of implementation. For example, recent literature reviews of MI in community (Mutschler et al., 2018) and school-based settings (Snape & Atkinson, 2016) found only 36% to 55% of published studies, respectively, reported fidelity data with the type and quality of data and procedural details reported varying widely.

Supporting Parents in the Context of School Mental Health

Parents play an important role in school mental health. Hoover-Dempsey et al. (2005) and Hornby (2011) have recognized the need for school mental health interventions to include a home module and attend carefully to parent engagement, motivation, and follow-through. Furthermore, Weist et al. (2014) identified family engagement as one of eight issues requiring systematic attention in order for the field of school mental health to advance. As well, Hoagwood and her associates (2007) noted effective school interventions for students requiring tertiary-level supports contain a well-designed and intensive family module that delivers the necessary intensity and dosage levels to substantively impact school outcomes and also address children's social, emotional and mental health difficulties.

Researchers and school-based personnel increasingly are using MI in school-based settings to engage and support parents either through brief, informal conversations (Rollnick et

al., 2016) or more structured intervention protocols (Sibley et al, 2016; Stormshak et al., 2020). In school-based settings, the use of MI with parents is particularly relevant given that: (1) parent practices supporting children's adjustment to the social and educational demands of schooling are positively associated with educational outcomes (Hoover-Dempsey et al., 2005; Huntsinger & Jose, 2009); (2) engaging parents can be challenging for school-based providers (Frey et al., 2013; Furlong & McGilloway, 2015; Hornby, 2011); and, (3) school-based mental health providers require efficient solutions such as MI given often cited barriers such as high caseloads and lack of time (Kelly et al., 2015; Thompson et al., 2019; Villarreal, 2018).

The homeBase (hB) intervention, the program for which the MI fidelity data reported in this paper was collected, targets the parents of children with early onset behavior problems. It can be delivered by school social workers or school-based mental health providers as an efficient supplement to school-based interventions (as was done in this study) or by community mental health providers as a brief, stand-alone home visitation intervention. The intervention is delivered via three to six visits with a student's parents. The sessions are designed to increase parent motivation and capacity to implement effective parenting practices. During hB sessions, the interventionist, hereafter referred to as a coach, uses MI to support parents as they reflect on and modify their parenting practices consistent with the five universal principles of positive behavior support (Sprague & Golly, 2013). The hB steps are engaging in values discovery (Step 1); assessing current parenting practices (Step 2); sharing performance feedback (Step 3); offering extended consultation and support (Step 4); and providing closure (Step 5). Additional details about hB can be found in Frey et al. (2019).

Sources of Variability in MI Fidelity

As Dunn et al. (2016) note, although reporting group-level means (e.g., average scores across all sessions) is commonplace, the procedure obscures meaningful forms of variability. There are different sources influencing variability in MI fidelity. These sources pertain to characteristics of: (a) the practitioner delivering MI; (b) the recipient of MI-infused services (e.g., a teacher, parent, or student in school-based contexts); and (c) the interaction between the two (Dunn et al., 20166). Imel et al. (2011) label this interactional variability between the practitioner and recipient of MI as "mutual influence," suggesting "client effects" and "relationship effects" are evident when variability in MI fidelity emerges within a therapist's caseload (e.g., MI quality varies by client). Hallgren et al. (2018) describe sources of variability in terms of provider-level, session-level, and site-level factors. Provider-level factors are equivalent to practitioner characteristics (e.g., education level, past exposure to MI training sessions) and session-level factors to the characteristics of those receiving MI services (e.g., severity, motivation for change). Site-level, contributory factors include variables such as organizational support for MI implementation, staff communication and cohesion, or access to resources that facilitate implementation and mitigate barriers (Hall et al., 2016).

There is ample evidence to suggest MI proficiency varies with respect to the aforementioned factors. Imel et al. (2011) found that therapists' MI skills were not consistent across clients served. In other words, therapists' competent delivery of MI varied within their caseloads. Specifically, they found that low client motivation at the outset of a session resulted in higher MI fidelity. Dunn et al. (2016) found (a) higher levels of variability within than between therapists and (b) stability in MI fidelity over time with scores neither significantly improving nor worsening over time. Finally, Hallgren et al. (2018) reported within-provider variability represented a much larger proportion of variance (i.e., between 57% and 94%) than between-

provider variability (i.e., between 3% and 26%). Thus, even highly competent providers may sometimes fail to implement MI with fidelity.

Below, we present group-level and coach-level MI fidelity data from a school-based efficacy trial and then examine between and within coach variability via partitioning of variance within three-level, multilevel models.

Method

Procedures for the Larger Efficacy Trial

Data for this paper are from the first four cohorts of an efficacy trial of the hB and First Step Next (FSN) interventions conducted in two Midwest school districts. FSN is a Tier 2 intervention targeting social emotional skills in the classroom (Walker et al., 2018). The study utilizes a 2 x 2 factorial design to examine the effect of FSN, the effect of hB, and the additive effect of the two interventions when delivered together. Children who met stage 2 criteria on the teacher-reported Systematic Screening for Behavior Disorders (Walker, Severson, & Feil, 2014) and were in the borderline or clinical range of the parent-reported CBCL's externalizing dimension (Achenbach, 2001) were eligible for participation. We recruited one student from each participating classroom and randomly assigned classrooms to one of four groups: hB only, FSN only, FSN plus hB, or control. The study was conducted in compliance with university and school district internal review boards. MI is part of the hB intervention but not the FSN intervention. Since the analysis reported herein is conducted on data collected during the MI-infused hB sessions, only the sample of participants randomized to either the hB-only condition or the FSN plus hB condition are included.

MI Training

We utilized the Motivational Interviewing Training and Assessment System (MITAS; Frey et al., 2017) to train the coaches participating in this study. The MITAS training model for this study included workshops; individualized coaching and feedback sessions; and monthly participation in a professional learning community (PLC), which are widely recognized as a common and necessary strategy for transitioning from training to skill maintenance among frontline professional and paraprofessional providers (Baez et al., 2020; Madson et al., 2016). Workshops introduced participants to the core elements of MI; facilitated development of the relational and technical components of MI; and helped promote skills needed to foster and encourage client-centered change talk. These workshops consisted of three, four-hour sessions. Participants then received three sessions of individualized coaching (a single participant completed only two coaching sessions due to scheduling complications). Coaching sessions ranged from 45 to 75 minutes. During these sessions, the participant – delivering hB as a behavioral coach – implemented each step of the program with an experienced coach who portrayed a "standardized parent." Finally, the research manager facilitated the MITAS PLC weekly, encouraging conversation and discussion among participants. During weekly PLC sessions, participants listened to and discussed audio-recorded conversations between coaches and parents and shared implementation successes and challenges. All training sessions took place at the University of Louisville.

Participants

One hundred sixty families were randomized to the hB-only or FSN plus hB conditions. Participating caregivers had a mean age of 35 years (*SD* = 9.4 years) and were predominantly female (88%). The majority reported their race as either African American (54%) or Caucasian (41%). Nine percent of parents held a bachelor's degree or higher. Nearly three-quarters of

parents were currently working (72%) and 33% were living below the poverty level based on reported income and household size. For 113 of the 124 participating families (91%) we had an audio recording of at least one hB session. There were no statistically significant differences in the characteristics of those with and without audio recordings.

All coaches were hired as employees of the University of Louisville to deliver MI as part of this project. The 20 participating coaches were primarily female (80%), and ranged in age from 23 to 61 years old (*M[SD]* = 33.6[12.8]). Seventy-five percent reported their race as white or Caucasian and 25% reported their race as black or African American. The coaches participating in this study had training, experiences, and credentials similar to school-based personnel and comparable to the interventionists delivering FSN in the classroom setting. Sixty percent held a Master's degree or higher. The remaining coaches were students pursuing a master's degree in social work. Ten coaches (50%) were trained as school social workers; three (15%) were trained as teachers; and one (5%) was trained as a school psychologist. The remaining six coaches (30%) were trained as community mental-health social workers. Participating coaches reported varied exposure to MI prior to training. Thirty percent had limited exposure and 25% had only read about the approach. The remaining 45% reported previously attending an MI training; though the duration and intensity of training varied.

Measures

We assessed MI proficiency using the MITI 4.2 (Moyers et al., 2015; Moyers et al., 2005). The MITI is a coding system used to examine the verbal behavior of a practitioner, counselor, or coach delivering MI. The MITI enables examination of the four MI processes of engaging, focusing, evoking, and planning through coding of four global scores and 10 behavior counts. A trained coder uses the MITI to review a random 20-minute audio segment, tallying

counts for each of ten behavior categories (e.g., simple reflections [SR], complex reflections [CR], affirmations, questions). Then, after listening to the audio segment, the coder provides a global rating on a 5-point scale for four global dimensions: cultivating change talk (CCT), softening sustain talk (SST), partnership, and empathy. The highest anchor for CCT indicates the coach or practitioner "shows a marked and consistent effort to increase the depth, strength, or momentum of the client's language in favor of change" (p. 5). The highest anchor for SST indicates "a marked and consistent effort to increase the depth, strength, or momentum of the client's language in favor of the status quo" (p. 7). These raw counts and scores are combined to generate four summary scores for (a) relational skills, (b) technical skills, (c) the percent of CRs, and (d) the ratio of reflections to questions. The relational global summary score is the mean rating of the partnership and empathy items. The technical global summary score is calculated as the mean score of CCT and SST. Percent of complex reflections is calculated by dividing CR by total reflections (e.g. SR + CR). Finally, as the name implies, the ratio of reflections to questions is the ratio of total reflections to the number of questions posed during a session.

Thresholds are based on expert opinion for basic and advanced fidelity[1]. For relational skills, scores greater than or equal to 3.5 indicate basic fidelity and scores greater than or equal to 4 indicate advanced fidelity. Thresholds for technical skills are scores greater than or equal to 3 (e.g., basic) and 4 (e.g., advanced). The percent of CRs above 40% indicates basic fidelity and above 50% indicates advanced fidelity. Finally, cutoffs for reflections-to-questions are a 1:1 ratio for basic fidelity and a 2:1 ratio or higher for advanced fidelity (Moyers et al, 2015).

Monitoring MI Fidelity

[1] The MITI manual refers to cutoff scores rather than thresholds and labels the minimum cutoff Basic Competency ("fair") and the advanced cutoff proficiency (good). We have changed the nomenclature in this manuscript to improve readability.

Coaches collected audio recordings of all hB sessions covering steps 1 through 3. We limited review of MI proficiency to these steps because they roughly align with the MI processes of engaging, focusing, and evoking change talk. Specifically, these three steps require more frequent use of MI as the coach develops a working relationship with the family (e.g., alliance); works with the participant to focus attention on areas for behavior change; and encourages talk about specific behavior change. Steps 4 and 5 were not coded given their focus, respectively, on skill building and closure. During steps 4 and 5, interventionists may use MI, but its use is not considered necessary for high fidelity implementation of the intervention. Session recordings for the three steps varied in length. On average, step 1 sessions lasted 59 minutes (*SD* = 19); whereas, step 2 (*M[SD]* = 55[33]) and step 3 (*M[SD]* = 47[22]) sessions were slightly shorter. At the end of each cohort, the third author (J.L.) prepared the digital audio recordings and provided them to an independent team of trained coders. The independent coders randomly selected a continuous 20-minute sample from each tape according to the project procedures detailed below and coded it using the MITI. Procedures for randomization varied by hB step to account for differences in the structure and timing of non-intervention, coach-caregiver interactions. For Step 1 recordings, coders extracted a random, 20-minute sample between the 20-minute mark of the recording and five minutes prior to the end of the recording. For Steps 2 and 3, coders randomly sampled a 20-minute segment between the 10-minute mark of the recording and five minutes prior to the end of the recording.

Coder Training

Three coders completed the MITI coding. All coders completed a two-day training on the MITI 4 and participated in ongoing group coding until reaching 90% reliability on behavior counts and 100% reliability on global scores. The last author of this publication (M.H.S.)

conducted IRR checks on a random sample of 20% of sessions. We assessed IRR via 2-way mixed effects, absolute agreement, average-measures ICCs. We used Cichetti & Sparrow's (1981) benchmarks to categorize the quality of the ICC. IRR was excellent for cultivating change talk (.777), partnership (.804), and empathy (.831) and – due, in part, to restricted range – was fair for softening sustain talk (.553). For technical global scores (.786) and relational global scores (.874) IRR was excellent. For percent CRs (.623) and reflections-to-question ratio (.703), reliability was good. In general, ICCs were comparable to those reported by the MITI developers (Moyers et al., 2016).

Statistical Analysis

To estimate the proportion of variance between coaches, families, and sessions, we fit unconditional three-level, random intercept models for each level-1 MITI summary measures. Although the MITI was used to code 249 sessions, only 245 sessions were included in the multilevel models. Four sessions were excluded to maintain balanced time across families (e.g., 1 to 3 sessions). For these families, two MITI observations were obtained for the same step because it was completed across two sessions. When this occurred, we used the first session.

Level-1 data consisted of MITI scores from 245 sessions collected on up to three occasions (*M[SD]* = 2.2[0.8]) per family. These data were then nested within level-2 families (n = 113), which were nested within level-3 coaches (n = 20). Models were fit in SPSS 24 using the restricted maximum likelihood estimator (REML). We compared nested models using the deviance difference in the -2 log likelihood (-2LL). For non-nested models, we examined values from the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). We calculated ICCs to assess the proportion of variance attributable to each level.

Results

In total, coaches delivered 77 step 1 sessions (31%); 89 step 2 sessions (36%) and 79 step 3 sessions (32%) across four cohorts. Mean scores for technical skills, complex reflections, and reflections-to-questions did not differ significantly by step. For relational skills, mean scores differed at step 1 (*M[SD]* = 4.0[0.7]) and step 3 (*M[SD]* = 3.5[0.8]; $F = 7.14$, $p < .001$). Coaches delivered between one and 42 sessions across one or more cohorts. Eight coaches who implemented across multiple cohorts accounted for 190 of the 245 sessions (78%). These coaches had an average of 23.8 sessions (*SD* = 11.5) whereas coaches who only worked for one cohort (n = 12) had an average of 4.6 sessions (SD = 3.3).

Group-level Fidelity

The top row in Table 1 summarizes mean scores across all 245 sessions for the four MITI summary scores. Across the 245 sessions, mean scores on the MITI global technical scale were in the basic fidelity range (e.g., ≥ 3.0). For all but seven sessions (97%), coach use of technical MI skills were above the basic fidelity threshold. On average, scores on the global relational scale were also in the basic fidelity range. For nearly 80% of sessions, global relational scores were above the basic fidelity threshold. For complex reflections and reflections-to-questions summary scores, 87% and 60% of sessions, respectively, exceeded basic fidelity thresholds. For 117 sessions, basic fidelity thresholds were met on all four MITI scores (48%). For 40 sessions, advanced fidelity thresholds were met across all four scores. For three sessions (1%), basic fidelity thresholds were not met for any of the MITI summary scores.

Coach-level Fidelity

Mean technical proficiency scores at the coach level ranged from 3.2 to 4.3; whereas mean scores for relational proficiency ranged from 2.7 to 4.4 (see Table 1). Average complex reflections by coach ranged from 33% to 77%. The reflections-to-questions ratio ranged from a

low of 0.1 (e.g., one reflection for every 10 questions) to a high of 3.4 (e.g., 3.4 reflections to each question). Across the four summary scores, coaches with more than 10 sessions of MITI data had mean scores comparable to coaches with fewer than 10 sessions of MITI data. With respect to session-level categorical cutoffs (e.g., all sessions above the specified cutoff), coaches with more than 10 sessions of MITI data were less likely to have all of their session above basic or advanced cutoffs as compared to coaches with fewer than 10 sessions of MITI data; though these differences were non-significant across all measures.

Table 2 aggregates the coach-level data reported in Table 1 to examine the number and percentage of coaches meeting basic proficiency and advanced proficiency cutoffs using (a) mean scores and (b) session-level categorical cutoffs. When using mean scores, the percent of coaches meeting basic proficiency ranged from 70% to 100% but, when applying session-level categorical cutoffs, the percent of coaches meeting basic proficiency dropped to a range of 20% to 70% depending on the summary measure. Similar drops occurred with respect to advanced proficiency as reported in the last two columns of Table 2, with advanced proficiency ranging from 0% to 15%. Eleven of 20 coaches (55%) met basic proficiency cutoffs across the four summary scores reported in Table 2. All 20 coaches met basic cutoffs on at least one score. Advanced proficiency cutoffs were more difficult to achieve even when using mean scores. Only two coaches (10%) had mean scores on all four summary measures exceeding advanced proficiency levels (one with 26 sessions of data and one with 15 sessions of data). Based on session-level categorical cutoffs, only two coaches (10%) met basic proficiency cutoffs across all summary scores (one coach with 15 sessions of data and one coach with a single session of data). Fifteen coaches (75%) met basic proficiency cutoffs on at least one summary score using

categorical cutoffs and five coaches (25%) met advanced proficiency using the categorical cutoff. No coaches met advanced proficiency across all sessions and all summary measures.

Between and Within-Coach Variability

To estimate the proportion of variance attributable to MI sessions (level 1), the families receiving hB support (level 2), and the coaches providing support (level 3), we fit unconditional three-level, random intercept models for each MITI summary score. For technical global, relational global, and reflections-to-questions summary scores, we tested whether a two-level model fit better than a three-level model. For technical global ($-2\Delta LL$ [4] = 25.6, $p < .001$) and relational global ($-2\Delta LL$ [4] = 37.3, $p < .001$), a three-level model fit better than a model with fewer parameters. For reflections-to-questions ($-2\Delta LL$ [4] = 2.6, $p = .108$), a two-level model fit better than a three-level model. The three-level model for percent of CRs did not converge. In turn, we fit and compared two, two-level models for percent of CRs, one nesting level-one variables within families and eliminating the coach-level and the other nesting level-one variables within coaches and eliminating the family-level. Based on a comparison of AIC and BIC values, the two-level model for CRs nesting sessions within coaches (AIC = 2240.4; BIC = 2247.4) fit better than the model nesting sessions within families (AIC = 2252.0; BIC = 2259.0).

Across all four models, between-session variability was the highest. As reported in Table 3, variance between sessions for the global and behavioral summary scores accounted for between 64% and 91% of variability. Variance between coaches accounted for 13% to 29% of variability. In contrast, between family variability accounted for only between 7% and 9% of variability. The ICCs indicate that between 30% and 37% of total variation in MITI technical and relational scores over time was attributable to variation at the family and coach level but that the

majority of this variation was attributable to the coach level. Specifically, between 77% and 79% of higher-level variability was attributable to the coach-level of the models.

Discussion

This study builds on the work of Dunn et al. (2016) and Hallgren et al. (2018) by examining between and within coach variability of MI fidelity within the context of an intervention to support parents of students at high risk for school failure. Similar to the aforementioned authors, we found MI quality varied between sessions and coaches and that within-coach variation (e.g., session-level variation in the quality of MI delivered) greatly exceeded between-coach variation. Specifically, the proportion of between-session variability (range: .64 to .91) was three or more times larger than the proportion of between-coach variability (range: .13 to .29). These findings suggest there is meaningful variation in coaches' MI skill (e.g., variability between coaches) but even greater variation in coaches' MI quality from session to session (e.g., variation between sessions). In addition to replicating previous finding in the context of SBMH services, this study provides a road map for other SBMH service practitioners and researchers to investigate MI fidelity. It is particularly important to do so with endogenous, school-based providers, and for MI-based interventions that are delivered in the context of the school—whether they focus on supporting teachers, parents, or adolescents.

Examination of proficiency through the lens of meeting or exceeding the MITI's basic fidelity thresholds tells a slightly different story, especially when considering mean coach-level scores. Applying cutoffs to coaches' mean scores on the four measures resulted in from 70% to 100% of coaches meeting basic proficiency on a given measure. These percentages dropped, however, when using a categorical cutoff requiring a coach to achieve basic proficiency *on every session*. For example, 70% of coaches had mean relational proficiency scores above the basic

cutoff but, when requiring a coach to meet the cutoff on all sessions, only 45% of coaches achieved this level of consistency. This shift is particularly noticeable with respect to advanced proficiency in the use of CRs and reflections-to-questions. Whereas coach-level mean cutoffs resulted in 80% of coaches reaching advanced proficiency in the use of CRs and 50% reaching advanced proficiency in reflections-to-questions, these percentages dropped to 20% and 0%, respectively, when requiring a coach to meet the advanced cutoff on every session.

As noted earlier, the number of hB sessions completed by a coach ranged from 1 to 40, depending largely on how long they worked on the project. Although this variation in the number of sessions may have made it easier (at least at first glance) for all of a coach's sessions to be above the categorical cutoff, there is ample evidence to suggest these cutoffs could be consistently met by some coaches as the number of sessions increased. For example, 5 of the 8 coaches with 10 or more sessions of data met the *basic proficiency* cutoff on all sessions of data for at least one summary measure. As well, one of the two coaches who met the *basic proficiency* categorical cutoff on all four measures had 15 sessions of data. As Table 1 suggests, consistency varied by measure and cutoff. Whereas coaches – regardless of the number of sessions delivered – were able to achieve consistent, *basic proficiency* for technical skills and CRs, consistently reaching *basic proficiency* for relational skills and reflections-to-questions was more difficult. For example, proficient use of reflections-to-questions increased, in general, with the number of sessions delivered. Seven of eight coaches with 10 or more sessions of data (88%) had mean reflections-to-questions proficiency scores above the *advanced* cutoff as compared to only three of 12 coaches with fewer than 10 sessions of data (25%). This suggests that proficient use of some skills may develop more slowly over time and with increased use of MI.

Practice Implications

In contrast to Hall et al.'s (2016) assertion that achieving MI proficiency may take years, our findings suggest it is possible to train coaches to basic proficiency levels within a few months; though reaching levels of consistent implementation of specific MI skills may take more time. Additionally, there is a need for efficient, cost-effective tools to measure MI fidelity in real-world contexts. Although collection and detailed examination of MI skills may be possible within efficacy trials, collection of fidelity data may be prohibitive in school settings due to a myriad of individual and contextual barriers. The Motivational Interviewing Evaluation Rubric, a recently published tool designed to increase MI implementation fidelity in community-based settings (Baez, et al., 2020), is one example of a new tool that may enable examination of MI proficiency and facilitate timely feedback to practitioners.

Future Research

Weisner and Satre (2016) encourage MI researchers to examine how much training is necessary to sustain MI fidelity over time; how it should be accomplished; and how the cost can justify the expense. We believe the MITAS is a good starting point for addressing these questions. With regard to future research, we encourage school-based researchers to replicate these findings. Additionally, we believe it would be beneficial to examine variation across program recipients (e.g., teacher, parent, or student) and to examine how data on satisfaction, alliance, and barriers collected after each session (e.g., time-varying covariates) affects variation in MI adherence and quality. The dosage and intensity of MI infused into school-based programs varies. Given the role this variation can potentially play in reducing the positive effects of MI (Miller & Rollnick, 2014), future research comparing fidelity across school-based MI programs targeting similar outcomes would be beneficial to future intervention development. As Miller and Rollnick (2014) discuss, MI training and fidelity have been linked to increased client change

talk and decreased sustain talk, which in turn predicts behavior change. Thus, another way to advance this line of research would be to examine fidelity within the context of the interaction between the coach and the recipient, as well as outcomes of interest.

Limitations

A notable limitation in our study is that we did not examine change talk or behavior change. It is critical to understand how MI fidelity relates to the recipient's talk about change and, ultimately, outcomes. Only within this broader context can the MITI's expert-driven MI fidelity thresholds be validated. As well, we did not collect session-level data on the coach or parent (e.g., alliance, satisfaction, etc.) or code session-level interaction data which, as noted early, could help inform our understanding of variation between sessions. Additionally, there are some limitations to current conceptualizations of MI fidelity. The MITI captures practitioner but not client verbal behavior (Jelsma et al., 2015). In turn, it measures the frequency but not the quality or sequencing of core communication skills (Moyers et al., 2015). As well, it was not developed specifically for use in school-based settings and does not examine differential use of MI skills across the four MI processes. Despite the availability of published proficiency thresholds, they are not empirically derived and, in turn, do not provide a clear indicator of proficiency (Miller & Rollnick, 2014). Finally, hB was designed as a cross-setting intervention to be implemented by school-based personnel in tandem with classroom support for the student. Given the development stage of this project (e.g., efficacy rather than effectiveness trial) we did not utilize endogenous providers (though we did recruit coaches with similar experience and qualifications) and were, therefore, unable to examine important organizational factors that could influence real-world implementation in school-based contexts.

Conclusion

Despite the growing use of MI in school-based settings, there is a dearth of articles examining proficient use of MI and describing practitioner- or coach-level variability. Detailed examination of variability is needed to inform training approaches for school-based personnel; to provide practitioners and researchers a roadmap to investigate MI competency and proficiency (as well as drift) in their own contexts; to better understand MI's unique impact within the context of efficacy and effectiveness studies; to provide context for feasibility studies examining the acceptability, demand, and practicality of the approach; and to enrich empirical investigations of MI implementation efforts in these settings. Thus, as researchers move beyond examining issues related to the uptake of MI in school-based settings, it is paramount that the proficiency levels of practitioners; the extent to which they adhere to core MI components; and how well they deliver MI to program recipients is prioritized and reported.

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflicts of interest.

Ethical approval: This study was approved by appropriate institutional and/or national research ethics committees (i.e., Oregon Research Institute, University of Louisville, and Jefferson County Public Schools). All procedures were performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in this study.

References

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. ASEBA.

Baez, J. C., Galanis, R., & Magill, M. (Online First). When the right measure doesn't exist: A novel motivational interviewing tool for community programs. *Child and Adolescent Social Work Journal*. doi:.org/10.1007/s10560-020-00650-y

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*(3-4), 31-34.

Dunn, C., Darnell, D., Atkins, D. C., Hallgren, K. A., Imel, Z. E., Bumgardner, K., … Roy-Byrne, P. (2016). Within-provider variability in Motivational Interviewing integrity for three years after MI training: Does time heal? *Journal of Substance Abuse Treatment, 65*, 74-82.

Frey, A.J., Lee, J., Small, J.W., Seeley, J.R., Walker, H.M., & Feil, E.G. (2013). Transporting motivational interviewing to school settings to improve engagement and fidelity of Tier 2 interventions. *Journal of Applied School Psychology*, *29*, 183-202.

Frey, A. J., Lee, J., Small, J. W., Walker, H. M., & Seeley, J. R. (2017). Motivational Interviewing Training and Support for School-Based Applications. *Emotional & Behavioral Disorders in Youth*, 17, 86-92.

Frey, A., Small, J. W., Lee, J., Crosby, S. D., Seeley, J. R., Forness, S., & Walker, H. M. (2019). homeBase: Participation, engagement, alliance, and social validity of a motivational parenting intervention. *National Association of Social Workers*, 1-8. doi: 10.1093/cs/cdz016

Furlong, M., & McGilloway, S. (2015). Barriers and facilitators to implementing evidence-based parenting programs in disadvantaged settings: A qualitative study. *Journal of Child & Family Studies, 24*, 1809-1818.

Hallgren, K. A., Dembe, A., Pace, B. T., Imel, Z. E., Lee, C. M., & Atkins, D. C. (2018). Variability in motivational interviewing adherence across sessions, providers, sites, and research contexts. *Journal of Substance Abuse Treatment, 84*, 30-41.

Hoagwood, K. E., Olin, S., Kerker, B. D., Kratochwill, T. R., Crowe, M., & Saka, N. (2007). Empirically based school interventions targeted at academic and mental health functioning. *Journal of Emotional and Behavioral Disorders, 15*, 66-92.

Hoover-Dempsey, K. V., Walker, J. M., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. (2005). Why do parents become involved? Research findings and implications. *The Elementary School Journal, 106,* 105-130.

Hoover, S. A. & Mayworm, A. M. (2017). The benefits of school mental health. In K. D. Michael & J. P. Jameson (Eds.) *Handbook of rural school mental health*. New York: Springer.

Hornby, G. (2011). *Parental involvement in childhood education: Building effective school-family partnerships.* New York: Springer.

Huntsinger, C. S. & Jose, P. E. (2009). Parental involvement in children's schooling: Different meanings in different cultures. *Early Childhood Research Quarterly, 24*, 398-410.

Imel, Z. E., Baer, J. S., Martino, S., Ball, S. A., and Carroll, K. M. (2011). Mutual influence in therapist competence and adherence to motivational enhancement therapy. *Drug and Alcohol Dependence, 115,* 229-236.

Jelsma, J. G. M., Mertens, V., Forsberg, L., & Forsberg, L. (2015). How to measure Motivational Interviewing fidelity in randomized controlled trials: Practical recommendations. *Contemporary Clinical Trials, 43*, 93-99.

Kelly, M.S., Thompson, Frey, A., A., Alvarez, & M., Berzin, S.C. (2015). State of School Social Work: Revisited. *School Mental Health.7,* 174-183 Doi: 10.1007/s12310-015-9149-9.

Lewis, C. C., Proctor, E. K., & Brownson, R. C. (2017). Measurement issues in dissemination and implementation research. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice*. Oxford University Press.

Macklem, G. L. (2014). *Preventive mental health at school: Evidence-based services for students*. New York: Springer.

Madson, M. B., Shumacker, J. A., Baer, J. S., & Martino, S. (2016). Motivational interviewing for substance use: Mapping out the next generation of research. *Journal of Substance Abuse Treatment, 65*, 1–5

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

Miller, W. R., & Rollnick, S. R. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: Impact of treatment fidelity. *Contemporary Clinical Trials, 37*, 234-241.

Moyers, T. B., Manuel, J. K., & Ernst, D. (2015). *Motivational Interviewing Treatment Integrity Coding Manual 4.2.1*.Unpublished manual.

Moyers, T. Martin, T., Manuel, J., Hendrickson, S., & Miller, W. (2005). Assessing competence in the use of motivational interviewing. *Journal of substance abuse treatment*, *28*, 19-26.

Mutschler, C., Naccarato, E., Rouse, J., Davey, C., & McShane, K. (2018). Realist-informed review of motivational interviewing for adolescent health behaviors. *Systematic Reviews, 7*, 1-21.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., …, Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Mental Health, 38*, 65-76.

Rollnick, S., Kaplan, S. G., & Rutschman, R. (2016). *Motivational interviewing in schools: Conversations to improve behavior and learning*. New York: Guilford Press.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 39*(4), 445-459.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460-475.

Sibley, M.H., Graziano, P.A., Kuriyan, A.B., Coxe, S., Pelham, W.E., Rodriguez, L.M. et al., (2016). Parent-Teen Behavior Therapy + Motivational Interviewing for Adolescents with ADHD. *Journal of Consulting & Clinical Psychology, 84*, 699-712.

Snape, L. & Atkinson, C. (2016). The evidence for student-focused motivational interviewing in educational settings: A review of the literature. *Advances in School Mental Health Promotion, 9*, 119-139.

Sprague, J. & Golly A. (2013). *BEST behavior. Building positive behavior support in schools.* 2nd ed. Longmont, CO: Sopris Learning.

Stormshak, E. A., McIntyre, L. L., Garbacz, S. A., & Kosty, D. B. (2020). Family-centered prevention to enhance parenting skills during the transition to elementary school: A randomized trial. *Journal of Family Psychology, 34*, 122-127.

Suldo, S. M., O'Brennan, L., Storey, E. D., & Shaunessy-Dedrick, E. (2018). Supporting high school students in accelerated courses. *Communique*, *46*(6), 1-21.

Thompson, A., Frey, A., & Kelly (2019). Factors affecting school social work practice: A latent profile analysis. School Mental Health, 11, 129-140.

Villarreal, V. (2018). Mental health referrals: A survey of practicing school psychologists. *School Psychology Forum, 12,* 66-77.

Walker, H., & Severson, H. H. & Fiel, E. G. (2014). *Systematic screening for behavior disorders (Revised 2nd Ed.): User's guide and technical manual.* Eugene, OR: Pacific Northwest Publishing.

Walker, H., Feil, E. G., Frey, A., Small, J., Seeley, J., Golly, A., … Stiller, B. (2018). First Step Next: An updated version of the First Step to Success early intervention program. *Perspectives on Early Childhood Psychology and Education, 3*, 89-110.

Weisner, C., & Satre, D. D. (2016). A key challenge for motivational interviewing: Training in clinical practice. *Addiction*, 111(7), 1154-1156.

Weist, M. D., Lever, N. A., Bradshaw, C. P., & Owens, J. S. (2014). Further advancing the field of school mental health. In M. D. Weist, N. A. Lever, C. P. Bradshaw, & J. S. Owens (Eds.), *Handbook of school mental health: Research, training, practice, and policy, 2nd ed*. New York: Springer.

Weist, M. D., Eber, L., Horner, R., Splett, J., Putnam, R., Barrett, S.,…, Hoover, S. (2018). Improving multitiered systems of support for students with 'internalizing' emotional/behavioral problems. *Journal of Positive Behavior Interventions, 20*, 172-184.

Table 1. Distribution of MITI proficiency scores, overall and by coach.

			Technical Proficiency			Relational Proficiency			Complex Reflections			R:Q Ratio		
			Scores	Global cutoffs		Scores	Global cutoffs		Scores	Global cutoffs		Scores	Global cutoffs	
Coach	# of families	# of sessions	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.
All	113	245	3.8 (0.5)	97	60	3.8 (0.8)	78	59	64.1 (24.0)	87	79	2.0 (1.9)	60	31
1	20	40	3.8 (0.4)	98	58	3.9 (0.5)	90	75	73.5 (14.8)	100	95	2.40 (2.49)	75	38
6	17	36	3.8 (0.5)	97	67	3.8 (0.7)	83	56	56.6 (23.1)	81	75	2.01 (1.96)	56	33
19	12	31	3.5 (0.4)	97	32	3.1 (0.8)	45	23	63.7 (27.8)	87	77	0.99 (0.83)	45	10
3	11	26	4.1 (0.3)	100	92	4.4 (0.4)	100	92	63.3 (21.4)	92	89	2.06 (2.48)	79	29
11	9	21	4.1 (0.4)	100	81	3.8 (0.8)	76	52	71.7 (16.6)	100	86	2.15 (1.61)	81	43
14	7	15	4.0 (0.4)	100	80	4.4 (0.4)	100	100	77.0 (17.9)	100	93	2.96 (2.10)	100	53
7	5	13	3.2 (0.5)	85	15	3.4 (0.7)	69	39	67.3 (30.7)	77	77	2.01 (1.34)	85	46
8	7	13	3.8 (0.3)	100	54	3.3 (0.8)	46	39	39.7 (19.9)	54	39	2.37 (1.47)	92	58
10	3	8	3.6 (0.4)	100	50	3.9 (0.4)	100	63	72.9 (27.5)	88	75	2.09 (1.08)	88	63
17	4	8	3.8 (0.7)	100	63	3.6 (0.8)	75	63	63.0 (13.3)	88	88	3.36 (2.99)	63	50

			Technical Proficiency			Relational Proficiency			Complex Reflections			R:Q Ratio		
			Scores	Global cutoffs		Scores	Global cutoffs		Scores	Global cutoffs		Scores	Global cutoffs	
Coach	# of families	# of sessions	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.	*M(SD)*	% above Basic	% above Adv.
2	3	6	3.3 (0.5)	83	17	2.7 (0.7)	17	0	57.8 (34.4)	83	83	0.79 (0.69)	40	0
13	2	6	3.5 (0.6)	83	50	3.9 (0.8)	83	67	62.0 (20.1)	67	67	0.95 (0.65)	33	17
16	2	4	3.9 (0.3)	100	75	4.1 (0.6)	100	75	72.8 (23.3)	100	75	1.74 (0.96)	67	33
9	2	4	4.0 (0.0)	100	100	4.0 (0.4)	100	75	46.0 (25.0)	75	25	1.55 (0.64)	100	33
15	2	3	3.8 (0.3)	100	67	3.8 (0.3)	100	67	60.0 (20.0)	100	67	2.39 (1.65)	67	67
5	2	3	3.5 (0.5)	100	33	3.2 (0.6)	67	0	33.3 (57.7)	33	33	0.10 (0.11)	0	0
20	1	3	3.8 (0.3)	100	67	4.0 (0.0)	100	100	75.9 (28.5)	100	67	0.94 (0.59)	67	0
4	2	2	4.3 (0.4)	100	100	4.3 (0.4)	100	100	60.0 (14.1)	100	100	0.56 (0.09)	0	0
12	1	2	3.8 (0.4)	100	50	3.3 (0.4)	50	0	35.0 (15.8)	50	0	1.86 (0.34)	100	50
18	1	1	3.5 (0.0)	100	0	3.5 (0.0)	100	0	66.7 (0.0)	100	100	1.00 (0.00)	100	0

Technical proficiency cutoffs: *Basic* ≥ 3.0, *Advanced* ≥ 4.0; Relational proficiency cutoffs: *Basic* ≥ 3.5, *Advanced* ≥ 4.0; Complex reflections cutoffs: *Basic* ≥ 40%, *Advanced* ≥ 50%; R:Q ratio cutoffs: *Basic* ≥ 1:1 (e.g., ≥ 1.0), *Advanced* ≥ 2:1 (e.g., ≥ 2.0).

Table 2. Number and percentage of coaches meeting basic and advanced proficiency cutoffs based on mean summary scores and session-level cutoffs.

	Basic proficiency		Advanced proficiency	
	Mean Cutoff *n(%)*	Categorical cutoff *n(%)*	Mean Cutoff *n(%)*	Categorical cutoff *n(%)*
Technical proficiency	20 (100.0)	14 (70.0)	5 (25.0)	2 (10.0)
Relational proficiency	14 (70.0)	9 (45.0)	6 (30.0)	3 (15.0)
Complex reflections	17 (85.0)	8 (40.0)	16 (80.0)	2 (10.0)
R:Q ratio	14 (70.0)	4 (20.0)	10 (50.0)	0 (0.0)

Table 3. Proportion of variance explained at session, family, and coach levels.

	Proportion of variance explained			ICCs	
MITI summary score	Between-session	Between-family	Between-coach	Level-2	Level-3
Technical global	.70	.07	.24	.30	.77
Relational global	.64	.08	.29	.37	.79
Percent complex reflections (%CR)	.88	--	.13	.13	--
Reflections-to-questions ratio (R:Q)	.91	.09	--	.09	--

Note: Level-2 ICCs = sum of between-family and between-coach variance; Level-3 = proportion of higher-level variability attributable to level-3 (e.g., between-coach/[between-family + between-coach]); "--" = Not estimated within best-fitting model.