Summer 6-1-2019

CLEARING THE MUDDY WATERS: IS LACK OF MOTIVATIONAL INTERVIEWING QUALITY ACCOUNTING FOR MIXED FINDINGS AND CAN MEASURING FIDELITY HELP?

Lauren N. Rowell
University of New Mexico - Main Campus

Lauren N. Rowell, M.S.
Candidate

Psychology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Theresa Moyers, Ph.D. , Chairperson

Jon Houck, Ph.D.

Kamilla Venner, Ph.D.

Tom Chavez, Ph.D.

CLEARING THE MUDDY WATERS: IS LACK OF MOTIVATIONAL INTERVIEWING QUALITY ACCOUNTING FOR MIXED FINDINGS AND CAN MEASURING FIDELITY HELP?

by

LAUREN N ROWELL

B.S., Psychology, University of North Carolina, 2009
M.S., Psychology, 2014

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Psychology

The University of New Mexico
Albuquerque, New Mexico

July, 2019

ACKNOWLEDGMENTS

I would like to thank my faculty mentor and thesis chair, Dr. Terri Moyers, not only for her time and extreme patience, but also for her intellectual contributions, guidance and support. Moreover, her steadfast belief in my ability as not only a clinician and scientist, but as a person, has been vital to shaping me into a psychologist. A thank you to my committee, Dr. Jon Houck, Dr. Kamila Venner, and Dr. Tom Chavez, for investing their time and energy in my training.

Also, this journey took a village. To my friends, for believing in me and encouraging me to reach for the stars. To my mother and sister, who have offered unwavering support through the ups and downs of this process. A heartfelt thank you to my father, for instilling in me the true value of education and teaching me that the sky is the limit.

And lastly, to my partner, for his sacrifices, understanding, and truly wanting, for us, whatever career path I want for myself.

CLEARING THE MUDDY WATERS: IS LACK OF MOTIVATIONAL INTERVIEWING QUALITY ACCOUNTING FOR MIXED FINDINGS AND CAN MEASURING FIDELITY HELP?

by

Lauren N Rowell

B.S., Psychology, University of North Carolina, 2009
M.S., Psychology, 2014
Ph.D., Psychology, 2019

ABSTRACT

Objective: The SMART ED study was a randomized, controlled trial of 3 brief interventions in ER settings targeting high-risk substance use. Findings in the main study indicated that there was no difference in outcomes for groups receiving minimal, rather than more extensive, behavioral interventions. This secondary analysis investigated the quality of (MI) sessions in the SMART ED study to examine the hypothesis that better quality of MI would be associated with improved client outcomes.

Method: The Motivational Interviewing Treatment Integrity Code (MITI 3.0) was used to evaluate 388 sessions, yielding indicators of MI Quality including Overall Spirit (MI Spirit), Reflection to Question (R:Q), Percent Open Questions (%OQ), MI Adherent (MIA) and MI Non-Adherent (MINA) behaviors. These quality indicators were used to predict client outcomes, measured as self-reported days of primary drug use, days with any drug use, and days of heavy drinking.

Results: Fewer than half (49%) of sessions met minimal criteria for MI quality using the MITI 3.0. Only 7% of sessions met the criteria for advanced MI practice. None of these indicators were associated with substance use outcomes.

Conclusions: These findings raise the possibility that negative trials of MI may be accounted for by poor adherence to the method, rather than lack of efficacy. Recommendations for quality monitoring and standards for future studies using behavioral treatments are discussed.

TABLE OF CONTENTS

LIST OF FIGURES ix

LIST OF TABLES x

CHAPTER 1 INTRODUCTION **1**

Motivational Interviewing and the Importance of Treatment Fidelity 1

Measuring Treatment Fidelity in MI Research 3

The Original Study: Background, Purpose, and Findings 5

The Current Study 5

Aim 1 6

Hypothesis Aim 1 6

Aim 2 6

Hypothesis Aim 2 7

Aim 3 7

Hypothesis Aim 3 7

Aim 4 8

Hypothesis Aim 4 8

CHAPTER 2 METHODS **9**

Participants 9

Measures 9

Interventions 10

Screening 10

Assessment 10

MI 10

Interventionists 11

Fidelity Measure: MITI 3.0 11

Sessions and Coders 11

Inter-rater Reliability 12

CHAPTER 3 RESULTS 13

Descriptive Statistics 13

Aim 1: Treatment Group Comparisons after Categorizing by MI Quality 13

1.1 Division of MI Treatment Group 13

1.2: Comparing Passable vs Not-Passable MI 14

1.3: Comparing Passable MI to the other Conditions 15

1.4: Comparing Beginner to Advanced MI 15

Aim 2: Data Driven and Theory Driven Analyses 16

2.1: Exploratory Factor Analysis 16

2.2: Data-Driven Regressions 17

2.3: Theory Driven Regressions 17

`Aim 3: Dose 17

` Aim 4: Therapist Effects: The proportion of outcomes accounted for by therapist 18

CHAPTER 4 DISCUSSION 19

Exploring Quality of MI as Related to Client Outcomes 21

Theoretically Driven Method 22

Empirically Driven Method 23

Dose 23

Therapist Effects 23

Summary and Implications 24

REFERENCES 33

LIST OF FIGURES

Figure 1. Percentage of the Sample that met MI Proficiency Standards 32

LIST OF TABLES

Table 1. Descriptions of MITI 3.0 Codes 26

Table 2. MITI Proficiency Standards..... 28

Table 3. Means and Standard Deviations of MITI Scores..... 29

Table 4. Percentage of Sessions that met each of the MITI 3.0 Proficiency Standards 30

Table 5. Obliquely rotated component loadings for the MITI 3.0..... 31

Chapter 1

Introduction

Motivational Interviewing and the Importance of Treatment Fidelity

Motivational Interviewing (MI) is a therapeutic method aimed to motivate clients towards health behavior change by exploring and resolving ambivalence (Burke et al., 2003; Miller & Rollnick, 2012). The modality incorporates both a relational component rooted in client-centered humanistic therapy (Rogers, 1951) and a technical component designed to facilitate client language in favor of change (Miller & Rose, 2009). Since the conception of MI in 1983, there has been extensive exploration into its efficacy for a variety of health behaviors including: alcohol and drug use, weight loss, diabetes management, adherence to medical treatments, sexual risk behaviors, gambling, and parenting behaviors (Armstrong et al., 2011; Harding et al., 2001; Parsons et al., 2005; Vasilaki, Hosier, & Cox, 2006; Webb, DeRubeis, & Barber, 2010) across a variety of settings including: medicine, counseling, and corrections (Lundahl et al., 2013; McMurran, 2009; Hettema, Steele, & Miller, 2005; VanBuskirk, & Wetherell, 2014). Recent MI meta-analyses find that 75% of participants show improvement in outcomes with small to moderate effect sizes compared to no-treatment conditions (Burke et al., 2003; Hettema, Steele, & Miller, 2005; Lundahl et al., 2013; Rubak et al., 2005; Valislaki, Hoiser, & Cox, 2006;). fMI improvements relative to other treatment modalities are mixed, with some meta-analyses finding equivalent outcomes (Hettema, Steele, & Miller, 2005; Lundahl et al., 2013) and others finding significant positive results with only small effect sizes (Rubak et al., 2005).

These modest effects may be accounted for, in part, by treatment fidelity. Fidelity establishes that the intervention given was the one intended (McLeod, Southam-Gerow, & Weisz 2009; Perepletchikova & Kazdin, 2005). It also allows potential differences in treatments to be observed if they exist. Further, measuring adherence offers researchers the opportunity to maintain the distinct elements of the intervention. This is important because one common issue in psychotherapy trials is drift: treatments wander from the what was originally planned (Beidas & Kendall, 2010; Chorpita & Nakamura, 2004; Damschroder & Hagedorn, 2011). Lack of attention to quality leads to little confidence in the interpretation of results for examinations of MI efficacy.

Fidelity problems are particularly critical for investigations of MI for a few reasons. First, a clear imbalance exists in quality of MI research. Some studies have been methodologically less rigorous than others with the gold standard being controlled clinical trials. (Burke et al., 2003; Miller & Moyers, 2015). Second, the training methods, supervision, and monitoring of therapists across randomized controlled trials (RCTs) greatly varies (Madson, et al. 2005). Third, there are concerns that some explorations have implemented MI in a way that violates its spirit (Moyers, Martin, Catley, Harris, & Ahluwalia, 2003; Rollnick & Miller, 1995).

Moreover, MI in its pure form is rarely used, even in pivotal projects such as MATCH. A vast majority of studies combine the method with other components and refer to this conglomerate as "MI" (Acosta, Haller, & Ingersoll, 2010; Madson et al., 2005; Miller & Sanchez, 1994; Miller & Rollnick, 2012; Kadden et al., 1998; COMBINE Study Research Group, 2003). Examples of this additive treatment include MI plus: 1) feedback (known as MET), 2) another type of treatment (such as Cognitive Behavioral

Therapy), and 3) psychoeducation (Miller, 1995; Parsons et al., 2005; D'Amico et al., 2015). This body of evidence, as it stands now, cannot elucidate whether these studies truly are engaging in the modality they claim. Furthermore, is it the MI, the additions, or the combination that are accounting for the current findings?

In addition to muddying the waters of research findings, lack of treatment fidelity measurement can be a hindrance to the *clinical* training and practice of MI. Treatment competence is a powerful clinical tool. It can be used to provide feedback which can help practitioners acquire new skills or continue to refine and improve techniques after learning them. Monitoring treatment quality can also help practitioners to better understand how their application of MI may change across different types of patient characteristics and populations. More importantly, MI integrity ensures best care and maximum impact of the treatment for clients (Burnam, Hepner, & Miranda 2009).

Measuring Treatment Fidelity in MI Research

Given the ways that a lack of fidelity can confuse the interpretation of RCTs and impact client outcomes, clinical scientists advocate for psychotherapy studies to consistently measure treatment quality (Southam-Gerow & McLeod, 2013; Perepletchikova & Kazdin, 2005). Despite its importance, most modalities have yet to create ways to measure intervention accuracy and only 3.5% of all treatment studies report adherence (Perepletchikova & Kazdin, 2005; Perepletchikova, 2011).

MI has an advantage in that many assessments have already been generated and evaluated. A recent systematic review found 5 measures that ranged in psychometric quality (Madson and Campbell, 2006). Although several excellent systems exist, the most

widely used is the Motivational Interviewing Treatment Integrity Code (MITI; Moyers et al., 2005; Moyers et al., 2016).

The MITI quantifies practitioners' ability to adhere to the theoretically proposed components of MI: the relational elements and technical skills. The global scores incorporate an overall impression of a therapeutic interaction, and are rated on a 5-point-Likert scale. On the other hand, the behavior counts involve a tally of specific verbal events across a session. The instrument also provides summary scores that are widely used in clinical settings to determine competency, but have yet to be empirically validated (Moyers et al., 2005). For a full description of all of these elements, see Table 1.

The MITI 3.0 has many strengths. The instrument is psychometrically sound having shown in multiple samples to have good reliability, validity, and sensitivity (Moyers et al., 2003; Moyers et al. 2005; Forsberg et al. 2007). Further, it is one of the few MI fidelity tools that is multimodal (global scores and behavior counts) and sets proficiency benchmarks, see Table 2 (Moyers et al., 2005). This measure is free of charge, has open source access, and is available at the following website: https://casaa.unm.edu/code/miti.html.

Surprisingly, given the availability and strengths of the MITI, treatment quality is variably reported in studies of MI efficacy (Handmaker, Miller, & Manicke, 1999; Madson & Campbell, 2006). As with other treatment types, MI researchers have suggested the need for future studies to measure and report treatment quality in order better understand the method's efficacy (Miller, 2001; Madson & Campbell, 2006).

Mixed findings in MI research can potentially be clarified by accounting for lack of treatment fidelity. One strategy to achieve this aim is to reanalyze data from existing studies that both reported quality *and* found no effect of MI on client outcome. In this study, a RCT was selected to conduct just such a secondary analysis.

The Original Study: Background, Purpose, and Findings

In order to ensure anonymity of the therapists, the study name and exact setting will be kept confidential. The parent project was conducted in a medical setting directly following urgent medical visits. The purpose was to investigate the effect of brief MI and referral for alcohol and drug users. The study compared the efficacy of three distinct treatment conditions: 1) minimal screening only (Screening), 2) screening, assessment, and referral to treatment (Assessment), and 3) screening, assessment, and referral plus brief MI and two follow-up MI booster sessions (MI). The authors hypothesized that the MI condition would decrease later alcohol and drug use as compared to the control conditions. Alcohol and drug use outcomes were: 1) self-reported days of primary drug use, 2) days with any drug use, and 3) days of heavy drinking, and were measured at 3, 6, and 12-months post treatment. The study yielded no significant differences between treatment conditions for alcohol and drug use across any follow-up point. Treatment fidelity was measured for the MI condition, using the MITI 3.0, but was only used to give clinical feedback to practitioners and was not reported in publications.

The Current Study

The purpose of this secondary analysis was to re-examine the MI treatment condition based on quality, as measured by the MITI benchmarks, to see if a significant treatment effect could be found. Given there has been debate over the meaning and use of

the terms “proficiency” and “competency” (Perepletchikova & Kazdin, 2005) this study will henceforth use the terms Beginner and Full-Strength MI, respectively. Our approach was multi-fold.

Aim 1. The first aim was to reanalyze the data, following the same analysis strategy as the original study, but using treatment quality as the grouping variable. To accomplish this aim, the baseline sessions of the MI treatment group were divided by quality, following the MITI competency benchmarks: Beginner versus Full-Strength. After this split, the differences between the two new MI conditions (Beginner and Full-Strength) on client outcomes (self-reported days of primary drug use, days with any drug use, and days of heavy drinking), were compared. Follow-up analyses investigated differences across all four treatment conditions: a) screening, b) assessment, c) Beginner MI and d) Full-Strength MI, using the same client outcome variables.

Hypothesis Aim 1. We expected that there would be a significant difference between MI groups, with the Full-Strength MI condition having better client outcomes as compared to the three other conditions (Beginner MI, Assessment, and Screening).

Aim 2. The second aim was to investigate if aspects of the MITI were associated with client outcome. In order to achieve this aim, two approaches were utilized: a data-driven and a theory-driven approach.

The data-driven approach determined the underlying latent variables for the MITI 3.0 and the factor loadings. Factor scores were computed for each therapy session. Then a regression was used to determine whether these factor scores accounted for a significant proportion of the variance in client outcomes.

The theory-driven approach examined whether particular constructs, important to MI theory, accounted for a significant proportion of the variance in client outcome. Given their theoretical significance and common clinical use, the MITI 3.0 summary scores (the average of the global scores, reflection to question ratio, percent open questions, percent complex reflections, and percent MI-Adherent behaviors) were investigated. We tested whether these variables accounted for a significant proportion of variance in client outcomes using regression analyses.

Hypothesis Aim 2. For the data-driven approach, we expected an underlying factor(s) would account for a significant proportion of the variance in client outcomes. For the theory-driven approach, we expected that all variables (an average of the global scores, reflection to question ratio, percent open questions, percent complex reflections, and percent MI-Adherent behaviors) would account for a significant proportion of the variance in client outcomes.

Aim 3. The third aim of this study was to ensure that the dose of MI was not conflating the competency findings in Aim 1. This was necessary as MITI scores used in the previous aims were only evaluated at baseline although many participants received two follow-up MI booster sessions. To accomplish this aim, we investigated whether dose, as measured by the number of sessions a participant received, was significantly associated with client outcome using linear regression.

Hypothesis for Aim 3. We expected that the dose of MI, as measured by the number of sessions a participant received, would not be significantly associated with client outcome.

Aim 4. The final aim of this study was to investigate how much variance in client outcomes was accounted for by the individual therapist.

Hypothesis for Aim 4. We expected that the particular therapist a client received would be associated with client outcome.

Chapter 2

Methods

Participants

Data were drawn from participants who were seeking medical treatment in six different locations across the United States. All were 18 or older, did not have significant impairment of cognition or judgment (i.e., delirium, traumatic brain injury or intoxication), spoke English, provided informed consent, had access to a telephone, indicated moderate to severe drug use problems, and reported at least 1 day of drug use in the 30 days prior to screening. Participants were compensated up to $275 for completing the study. For the current project, all participants in the three treatment conditions were included (N=1246) with the exception of 34 participants from the MI condition who were missing audio recordings.

Measures

All participants received a baseline assessment before randomization to treatment groups. To assess for alcohol and drug consumption, the following measures were included the: 1) Thirty-day timeline follow-back (TLFB) interview (Sobell & Sobell, 1992) and 2) NIDA-Modified version (NMASSIST) of the WHO ASSIST (Humeniuk, Ali, & Babor, 2008). Substance use related problems were quantified by the following: 1) Drug Abuse Screening Test-10 (DAST-10; Skinner, 1982) with scores indicating drug problem level (0= none, 1-2= Low, 3-5= Moderate, 6-8= Substantial, 9-10= Severe), and 2) Alcohol Use Disorders Identification Test (AUDIT-C; Bush et al., 1998) with scores indicating hazardous drinking (men scoring 4 or more and women scoring three or more). The follow-up measures (self-reported days of primary drug use, days with any drug use,

and days of heavy drinking) were analyzed at three time points: 3, 6, and 12-months post-treatment as measured by the TLFB interview. This secondary analysis used the same outcome measures as the original study for consistency in comparing results. However, only outcomes at one time point, three months, will be investigated in order to reduce the total number of analyses. This reduction both increased power and decreased inflation bias (Head et al., 2015).

Interventions. Participants in the original study were randomized to one of three groups: Screening, Assessment, and MI. All interventions (minus the MI booster sessions) were delivered around each participant's urgent medical visit.

a) Screening. Participants were given an informational pamphlet about drug use and its consequences written by the National Institute on Drug Abuse.

b) Assessment. Participants received the same pamphlet as in the screening group. Additionally, participants were given scripted feedback and referral if warranted. Feedback included personalized information aimed at warning participants about the consequences of their high-risk substance use. Referral included a standardized list of local treatment agencies.

c) MI. Participants received the same pamphlet, scripted feedback, and referral as in the Assessment condition. Furthermore, the participants received a 30-minute manual-guided MI session. The intervention consisted of a personalized feedback report (which included quantity/frequency of use, monetary consequences, risk factors for dependence, and normative feedback) and the development of a change plan (which included a discussion about hypothetical or actual change and developing a plan for change, if clinically appropriate).

Between seven days and one month after the initial session, participants received two twenty-minute follow-up MI booster phone calls. These check-ins included a discussion aimed at revising the participant's change plan, reinforcing and supporting change efforts, and exploring barriers to treatment. Due to attrition, less than half of the participants received the follow-up booster sessions.

Interventionists. Forty-two research staff, who did not work at the medical center and had no prior clinical experience, were hired to administer the interventions. To control for potential therapist effects all interventionists administered all three treatment conditions.

Training of the therapists included a two-day MI workshop (including lecture, role play and practice), an additional two-day study-specific training (unspecified), and feedback from two practice sessions for which the study supervisor (an expert in MI) measured treatment fidelity. Throughout the study, interventionists had regular supervision which included MITI 3 coding of one of their recent session recordings.

Fidelity Measure: MITI 3.0. The MITI 3.0 was used to evaluate all baseline MI sessions by the lead supervisor. Definitions of the global scores, behavior counts, and summary scores for this measure are found in Table 1. Competency was defined by the following MITI 3.0 guidelines, see Table 2.

Sessions and Coders. All of the MI intervention sessions from the parent project (N=388) were included in the current secondary analysis. Three raters were trained in the MITI 3.0**.** Coders were deemed competent to begin coding data for the parent study after receiving good to excellent reliability on all items when compared to an expert coder.

Weekly coding meetings were held in order to resolve discrepancies and to prevent coder drift.

Inter-rater Reliability (IRR). A random twenty percent of the sessions were double coded by an expert rater and used to estimate IRR. Following recent recommendations (Hallgren, 2012). IRR was assessed with 2-way mixed effects, absolute agreement, and single measures intraclass correlations (ICC) for each individual item. The conventional benchmarks for ICCs are as follows: 0.00–0.40 = poor, 0.40–0.59 = fair, 0.60–0.74 = good, and 0.75–1.00 = excellent (Cicchetti & Sparrow, 1981). All global and behavior count scores were in the fair to excellent range.

Chapter 3

Results

Descriptive Statistics

Participants were 70% male with a mean age of 36 years. The ethnicity of the sample included: 50% White, 20% African American, 28% Hispanic, and 2% other. The participants' drug of choice included: 45% Cannabis, 28% Cocaine, 18% Street Opioids, 5% Prescription Opioids, and 4% Methamphetamine. The average score on the DAST-10 was 5.77 (SD = 2.28) indicating a substantial level of drug related problems and the range was 3-10 indicating moderate to severe problems. The average of the AUDIT-C was 5.43 (3.79) indicating hazardous drinking. Baseline averages for the three outcome measures were: 1) days of primary drug use (M=15.88, SD=8.70), 2) days of heavy drinking (M=4.06, SD=6.75), and 3) days of any drug use (M=17.52, SD=8.42).

Aim 1: Treatment Group Comparisons after Categorizing by MI Quality

1.1: Division of the MI Treatment Group. Baseline MI sessions (N=388) were divided into: Beginner and Full-Strength using the MITI 3.0 competency standards (Table 2). Descriptive statistics of each of the MITI summary scores for this sample are summarized in Table 3.

For participants in the parent study MI condition, 51.8% did not meet minimal standards, 41.5% met the threshold for Beginner MI, and 6.7% met the threshold for Full Strength MI (Figure 1). Descriptive statistics showing the percentage of intervention sessions that met each of five MI fidelity criteria are reported in Table 4. The range and mean of clinician scores for each of the MITI criteria are visually depicted via boxplots in Figures 2-5.

To compare the Passable and non-Passable MI conditions with respect to the outcome variables, three ANOVAs were computed using a linear mixed model, controlling for treatment site as a fixed effect and baseline use as a covariate. No significant differences between MI groups for days of primary drug use ($F(1, 12) = 97.148$, $p = .000$), days of heavy drinking ($F(1, 12) = 125.502$, $p = .000$), and days of any drug use ($F(1, 12) = 77.836$, $p = .000$) existed although there were significant differences in baseline use. The effect sizes, partial eta squared, were .230, .278, .193, respectively.

1.2: Comparing Passable vs Not-Passable MI on Major Study Outcomes.

The distribution of the three outcome variables were examined for skewness and kurtosis. They were zero-inflated, but otherwise had normal distribution. Following recent guidelines for zero-inflated data, no transformations were used because less than 20% of each outcome variable was zeros (Hallgren & Witkiewitz, 2014).

T-tests were conducted to examine if the treatment groups differed in baseline use. There were no significant differences between the Passable MI condition and Not-Passable MI condition for baseline days of primary drug use ($t(381) = -.394$, $p= .239$), days of heavy drinking ($t(381) = .139$, $p = .551$), and days of any drug use ($t(381) = .097$, $p = .233$). Additionally there were no significant differences between the two MI groups and the assessment condition for baseline use on all three variables respectively ($t(812) = -1.759$, $p= .283$; $t(812) = .876$, $p = .368$; $t(812) = -2.792$, $p = .922$). The screening condition did not collect these variables and therefore they could not be included. Despite the absence of significant differences, analyses included baseline use as a covariate because it has been found to independently predict follow-up use.

To compare the Passable and non-Passable MI conditions with respect to the outcome variables, three ANOVAs were computed using a linear mixed model, controlling for treatment site as a fixed effect and baseline use as a covariate. No significant differences between MI groups for days of primary drug use ($F(1, 12) = 97.148$, $p = .000$), days of heavy drinking ($F(1, 12) = 125.502$, $p = .000$), and days of any drug use ($F(1, 12) = 77.836$, $p = .000$) were found although there were significant differences in baseline use between groups. The effect sizes, partial eta squared, were .230, .278, .193, respectively.

1.3: Comparing Passable MI to the Screening and Assessment Conditions.

The original study analysis was replicated using a mixed linear model with treatment group (Passable MI, Screening, Assessment), and treatment site as fixed factors and baseline use as a covariate. No significant differences were found between the three treatment groups for days of primary drug use ($F(1, 12) = 133.656$, $p = .000$), days of heavy drinking ($F(1, 12) = 133.555$, $p = .000$), and days of any drug use ($F(1, 12) = 117.781$, $p = .000$) although the groups significantly differed in regards to baseline use. The effect sizes, partial eta squared, were .130, .130, .117, respectively.

The screening group's baseline use was estimated from an average of the two other conditions (MI and Assessment) for this analysis because this data was not collected.

1.4: Comparing Beginner to the Advanced MI.

To compare the Beginner and Advanced MI conditions with respect to the outcome variables, three ANOVAs were computed using a linear mixed model, controlling for treatment site as a fixed effect and baseline use as a covariate. No

significant differences between MI groups for days of primary drug use ($F(1, 6) = 99.609$, $p = .000$), days of heavy drinking ($F(1, 6) = 131.491$, $p = .000$), and days of any drug use ($F(1, 6) = 78..958$, $p = .000$) were found although there were significant differences in baseline use between groups. The effect sizes, partial eta squared, were .239, .281, .190, respectively.

Aim 2: Data Driven and Theory Driven Analyses

2.1: Exploratory Factor Analysis (EFA). An exploratory factor analysis was conducted using Principal Axis Factoring and Orthogonal Varimax Rotation to derive estimates of underlying factors within the MITI 3.0. An examination of the Kaiser-Meyer Olkin Test for sampling adequacy suggested that the data were factorable. The Bartlett's Test of Sphericity showed patterned relationships for the individual items of the measure. Three factors had eigenvalues over 1, and were retained. They explained 65.67% of the variance. Table 5 shows the factor loadings after rotation using a significant factor criterion of 0.4.

Seven items: all five global scores (Collaboration, Autonomy Support, Empathy, Evocation, and Direction) and Complex Reflections loaded onto a first factor. This was labelled "MI Spirit" as the included components mostly relate to the relational aspects of the method. The MITI heavily relies on Complex Reflections in the operationalized definition of Empathy so it follows that these scores were included together.

Six items: MI-Adherent and MI-Non-Adherent, Closed Questions, Simple and Complex Reflections, Giving Information, all behavior count scores, loaded onto a second factor. This was labeled "Behavior Counts" because these components were all measured as tally scores and don't otherwise relate to each other theoretically.

Three items (Evocation, Direction, and Open Questions) loaded onto the third factor. This was labelled "Evocation" because these components relate to the ability to move the client towards change in a collaborative way.

Factor scores were then calculated.

2.2: Data-Driven Regressions. Three regression analyses were conducted for the MI group with factor scores as predictor variables, baseline use as a covariate, and the three outcomes as the dependent variables. None of the factor scores were found to significantly affect any of the three outcomes and the only significant predictor was baseline use [$F(1, 12) = 94.007$, $p = .000$; $F(1, 12) = 140.576$, $p = .000$; $F(1, 12) = 76.960$, $p = .000$].

2.3: Theory Driven Regressions. Three stepwise multiple regressions were conducted using only the Passable MI group. Predictor variables in these models were an average of the global scores, reflection to question ratio, percent open questions, percent complex reflections, and percent MI-Adherent. Baseline substance use was a covariate. None of the predictor variables were found to significantly affect any of the three outcomes and the only significant predictor was baseline use [$F(1, 12) = 57.397$, $p = .000$; $F(1, 12) = 66.900$, $p = .000$; $F(1, 12) = 41.667$, $p = .000$].

Aim 3: Dose

Linear regressions, using only the Passable MI group, with dose, as measured by the number of sessions (one, two, or three) a participant received as a predictor variable, with baseline use as a covariate, and each of the three-month outcomes as the dependent variables were computed. Dose was not significantly related to three-month days of primary drug use, days of heavy drinking, or days of any drug use, although baseline use

was significant for each of the models [$F(1, 12) = 57.397$, $p = .000$; $F(1, 12) = 66.900$, $p = .000$; $F(1, 12) = 41.667$, $p = .000$].

Aim 4: Therapist Effects: The proportion of outcomes accounted for by therapist.

Data were analyzed using hierarchical linear modeling to assess the proportion of days of primary drug use, days of heavy drinking, or days of any drug use at 3-month post treatment that were accounted for by individual interventionist. We used an unconditional two-level model with clients nested within therapists. We found that 2.2% of the variance in days of heavy drinking, 1.17% of days of any drug use, and 0.004% of days of primary drug of choice was explained by the therapist.

Chapter 4

Discussion

One clear take-away from these data is that most participants in the parent study did not receive the intended intervention. The majority of therapists did not meet minimum standards for MI competency as defined by the MITI. Additionally, of those that did, they only met beginner not advanced standards.

Given so few sessions met MI quality, it is helpful to discuss how clinicians excelled and fell short. The clinicians were successful in many ways: they consistently met the bar for global ratings, such as empathy and partnership, the number of complex as opposed to simple reflections, and behaviors consistent with MI. In contrast, clinicians rarely met two criteria to be considered proficient in MI: 1) asking more open than closed questions and 2) making more reflections than questions. It may be that the latter two skills are harder to learn for beginning clinicians. This study shines light on particular abilities that may warrant increased attention for training therapists in both research and clinical work.

In some ways, the inability of these clinicians to meet MITI thresholds, is not surprising. It is possible that these interventionists may not have received sufficient training to learn the method or to become advanced in it. Although their training included a MI and study-specific workshop (both two days in length), as well as personal feedback from two practice sessions, this may haven fall short. In one seminal randomized controlled trial of methods of training MI, the Evaluating Methods for Motivational Enhancement Education (EMMEE) study, a two-day training was not sufficient to allow therapists to acquire proficiency one year later (Miller et al., 2004). Rather the whole

training package (i.e. a two-day workshop, two personal feedback reports on MI performance, and six individual coaching) was associated with better skill acquisition and more changes in client language (more change talk and less sustain talk) than any of these constituent parts alone. The current study included less coaching and feedback over time. This amount of continued practice and instruction may have been insufficient to obtain adeptness in the method. The current study further suggests, just like the EMMEE study, that individualized attention over time is helpful to gain expertise. No one presumes a novice pianist can take a two-day workshop and then be able to play a masterpiece, yet this was expected of these novice clinicians learning a complex therapeutic method. This phenomenon is not exclusive to learning a new therapy modality, but relevant to acquiring any ability. Evidence of this concept has been supported in the task performance and learning literature with a meta-analysis indicating spaced practice is superior to massed practice (Donovan & Radosevich, 1999). Our findings highlight the importance of training over time for acquiring therapeutic skill.

Another reason these clinicians may not have met acceptable standards of MI practice is that these standards are unrealistic. MITI summary scores have been routinely hard to achieve in large scale MI training studies. Despite the 85% proficiency rate attained in the EMMEE trial, others have been unable to reach such a high benchmark. In three such studies, this rate ranged from 4.3% to 36% (Baer et al., 2004; Moyers et al. 2008; Tollison et al., 2008). Moreover, just like our secondary analysis, clinicians were unable to meet two specific criteria: 1) asking more open than closed questions and 2) making more reflections than questions.

In fact, only 16% or fewer interventionists in these training trials met competency on these two indices. This lends further standing to our findings, which suggest these specific indicators may be impractical.

Despite this level of training, less experienced clinicians (as in the current investigation) may present a particular challenge. For example, in the Moyers et al. (2008) training study, the clinicians had less education and fewer years of counseling experience than the EMMEE trial and had a proficiency rate that ranged from 4.3%-10.5%, depending on training enrichments. They still had on average 11 years of counseling involvement, while therapists in our study had no clinical experience at all. If current MITI criteria are difficult to achieve with more seasoned clinicians, it may be impractical to expect as much of novices with no experience. This presents a dilemma for RCTs: to hire more experienced, more expensive clinicians or to hire less experienced, cheaper clinicians. A common way to save money in the short run is to hire beginners, but in the long run this may be costlier if time and money is not spent to train them to competence. Our findings lead further support to the recommendations set forth by other psychotherapy trials which suggest providing high quality training over time, continued monitoring of clinicians throughout a trial, and dismissal of therapists that do not meet standards (Miller, 2001; Perepletchikova & Kazdin, 2005).

Exploring Quality of MI as Related to Client Outcomes

An over-arching postulation made during this secondary analysis was that the quality of the intervention would predict client substance use outcome, but we found no support for this, despite employing multiple strategies for analyses.

Theoretically Driven Method. Our hypothesis was not supported; none of the MITI 3.0 summary scores (the averages of the global scores, R:Q, %OQ, %CR, and %MIA) were associated with better outcomes despite the level of clinician skill. This is contrary to a substantial body of research that suggests one proposed active ingredient of MI is change talk, or client language for change (Miller & Rose, 2009; Vader et al., 2010; Amrhein et al., 2003). Specifically, several studies suggest change talk is associated with better client outcomes and that this relationship is mediated by therapists' in-session behavior (Aharonovich et al., 2008; Baer et al., 2008; Hodgins, Ching, & McEwan, 2009).

The clinician summary variables we explored have been consistently shown in the literature to be associated with client change language in MI sessions. For example, global scores are positively associated with alliance and engagement as well as higher levels of client change talk (Boardman, Catley, Grobe, Little, Ahluwalia, 2006). Further, MI-Adherent behaviors are associated with both more client language for change in session (Gaume et al, 2010; Glynn & Moyers, 2010;). Given this array of evidence, our inability to find any aspects of the MITI associated with client outcome should be taken with a grain of salt.

It is possible that we were unable to detect these associations because of the restricted range of our data. Future studies examining similar questions might aim for a more diverse sample, particularly ones that meet adherence. It is also plausible that the fidelity measurement tool itself needs refinement. A weakness of the current MITI standards is that they are derived from expert opinion. Additionally, the MITI was crafted

for differentiating other modalities from MI and for recognizing beginner level MI, so it is possible the instrument is not as suitable for discriminating novices from experts.

Empirically Driven Method. The EFA yielded non-intuitive results which did not predict client outcomes. The factors that emerged seem to have clustered together due to similarities in the method of measurement rather than similarities in their constructs. This seems likely because the first factor included predominantly items measured by a Likert-scale and the second factor included only items measured by tally scores. As well, the original MITI was derived using an EFA of a more in-depth measure, the Motivational Interviewing Skill Code (Moyers et al., 2005), so conducting this analysis may not have been warranted. Something already in its simplest form cannot be further simplified.

Dose

The number of MI sessions each participant received (one, two, or three) was not associated with outcome. This finding seems intuitive as there is no reason to expect dose to be associated with better outcomes when participants did not receive the intended intervention.

Therapist Effects

The final hypothesis was that the therapist a client received would significantly impact their drinking and drug use. We found that therapists accounted for 2% or less of the variance in participant outcomes despite being rated well on global ratings such as partnership and empathy. This finding is unlike other recent substance abuse research where therapist effects have been more substantial (Miller and Moyers, 2015). In three such studies, these effects ranged from 11% and 67% (Moyers, Houck, Rice,

Longabaugh, & Miller, 2016; Miller and Moyers, 2015; Miller, Taylor, & West, 1980). Also, our findings are contrary to the body of evidence that clinician interpersonal skill and ability to create an alliance with the client is correlated with larger differences in outcomes than the differences produced by the individual treatment modality (Miller and Moyers, 2015; Messer & Wamphold, 2002).

One explanation for our null findings could be that global score measurement is particularly unreliable and less rigorous as opposed to a tally score. Recent research has highlighted concerns with Likert-scales in social sciences. One limitation is that respondents are artificially required to pick a stance (even if they lack an opinion), because these scales have no neutral point (Brown, 2006; Croasmun & Ostrom, 2011). Further, in Likert-scales with odd numbers, such as the MITI, the middle option acts as "neutral" and is more frequently chosen in moments of indecision, which increases response bias (Croasmun & Ostrom, 2011; Fernandez & Randall, 1991). Another possible explanation for the lack of therapist effects in this sample is that interventionists were crossed with treatment conditions. Because the therapists were evaluated highly on their global ratings, which encompass common factors such as empathy and partnership, an interventionists interpersonal skill may have potentially bled through into the screening and assessment conditions. This hypothesis is further supported by the fact that in the parent study all treatment conditions had improvements in drinking and drug use at 3 months after baseline. Given these shortcomings, a more concrete form of measuring clinician global scores may be warranted in future studies.

Summary and Implications

Overall, the conclusion from this secondary analysis is clear: MI was not effective in reducing substance use with clients present for urgent medical care. These data raise the hypothesis that MI might not be applicable in this hectic environment. It may be that individuals presenting for a pressing physical health issue are not as ambivalent about their substance use. This may be particularly true as they were non-treatment seeking for substance use issues. In support of this inference, four other trials researching the efficacy of MI in this real-world environment found low effect sizes ranging from .02-.16 (Vasilaki, Hoiser, & Cox, 2006; Gentilello et al., 1999; Longabaugh et al., 1996; Monti et al., 1999). Given the minimal impact of this treatment urgent care, future research should focus on methods that can more easily translate to patient care.

One final deduction from these data is that monitoring and reporting treatment quality are essential to efficacy research. As this study shows, in order to be sure (or in this case not sure) that the intervention given was the one intended it is *essential* to measure and report fidelity. Measuring competency is time-consuming, tedious, and expensive but vital. A final recommendation is for RCTs to assess integrity continuously, not just as an afterthought. Our secondary analysis was unable to paint a clearer picture of how MI quality is associated with client outcome, but has highlighted the importance of trying to clear up these muddy waters by placing a higher value on treatment fidelity.

Appendix A. Tables

Table 1. *Descriptions of MITI 3.0 Codes*

MITI Code	Brief Description
Globals	
Evocation	Conveys an understanding that motivation and ability for change reside within the client.
Collaboration	Encourages power sharing with client and allows client's ideas to influence the session.
Autonomy/Support	Supports and fosters client perception of choice.
Direction	Maintains focus on a specific target goal.
Empathy	Understands or makes an effort to grasp the client's perspective and experience.
Behavior Counts	
Giving Information	Gives information, educates, provides feedback, or expresses a professional opinion without persuading, advising, or warning.
MI Adherent	Category that includes behaviors consistent with MI spirit including: asking permission before giving advice, saying something positive or complimentary of a client, emphasizing the client's autonomy, and statements of compassion or sympathy.
MI Non-adherent	Category that includes behaviors inconsistent with MI spirit including: advise without permission, confronting, or giving orders or commands.
Open Question	A question with a wide range of possible answers.
Closed Question	A question that can be answered with a "yes" or "no" response or with a restricted range.
Simple Reflection	Reflects a client's statement with little or no added meaning or emphasis.
Complex Reflection	Reflects a client's statement with added meaning or emphasis.
Summary measures	
Global Spirit Rating	(Evocation + Collaboration + Autonomy/Support + Direction + Empathy)/ 5
Percent Complex Reflection	Complex Reflections/(Simple Reflections + Complex Reflections)
Percent Open Questions	Open Questions/(Closed Questions + Open Questions)

Reflection to Question Ratio	(Simple Reflections + Complex Reflections)/ (Closed Questions + Open Questions)
Percent MI Adherent	MI Adherent/ (MI Adherent + MI Non-Adherent)

Notes. Table created from the manual by Moyers, Martin, Manuel, Miller, and Ernst (2007).

Table 2. *MITI Proficiency Standards*

MI Summary Score Thresholds	Beginner	Full Strength
Global Clinician Ratings	Mean of 3.5	Mean of 4
Reflection to Question Ratio (R:Q)	1	2
Percent Open Questions (%OC)	50%	70%
Percent Complex Reflections (%CR)	40%	50%
Percent MI-Adherent (%MIA)	90%	100%

Table 3. *Means and Standard Deviations of MITI Scores*

MITI Scores	Beginner MI (N=159)		Advanced MI (N=27)	
	M	SD	M	SD
Global Scores				
Evocation	4.52	0.63	4.62	0.50
Collaboration	4.57	0.58	4.62	0.50
Autonomy Support	4.46	0.55	4.46	0.51
Direction	4.83	0.43	4.88	0.33
Empathy	4.66	0.51	4.65	0.49
Behavior Counts				
Giving Information	1.94	2.17	1.08	1.85
MI-Adherent	4.69	3.22	4.31	2.96
MI-Not adherent	0.02	0.13	0.00	0.00
Closed Question	4.73	3.00	2.58	1.90
Open Question	9.29	4.76	8.19	4.40
Simple Reflection	8.10	4.90	8.88	5.84
Complex Reflection	15.69	6.21	18.73	8.29
Summary Scores				
Average of the Global Scores	4.60	0.44	4.65	0.35
R:Q	2.02	1.25	3.25	2.17
%OC	66.78	12.01	79.19	10.13
%CR	66.80	13.19	68.50	10.86
%MIA	99.84	1.20	100.00	0.00

Table 4. *Percentage of Sessions that met each of the MITI 3.0 Proficiency Standards*

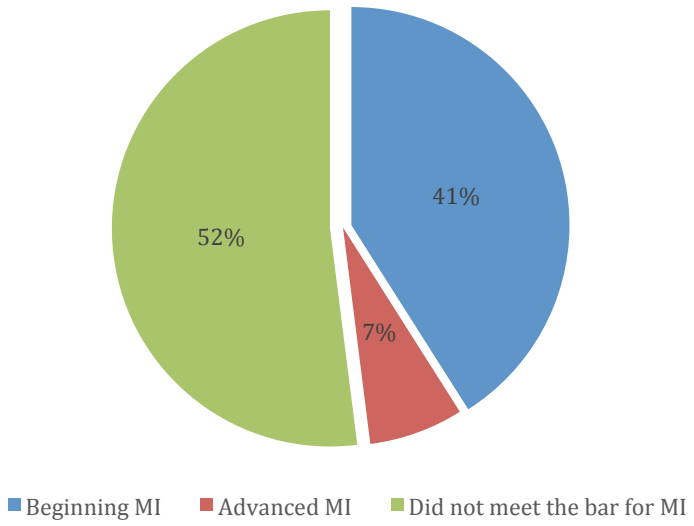
Clinician Summary Score Thresholds	Percentage of Sessions that met the Standard	
	Beginner MI	Full Strength MI
Global Clinician Ratings	93.0%	81.96%
Reflection to Question Ratio (R:Q)	74.2%	25.3%
Percent Open Questions (%OC)	58.5%	22.9%
Percent Complex Reflections (%CR)	91.2%	78.4%
Percent MI-Adherent (%MIA)	90.6%	85.8%
Met all 5 Threshold Scores	41.5%	07.3%

Table 5. *Obliquely rotated component loadings for the MITI 3.0*

Component	Factor 1: MI Spirit	Factor 2: Technical	Factor 3: Evocation
Collaboration	.872		
Autonomy Support	.860		
Empathy	.821		
Evocation	.806		.321
Direction	.534		.329
MI-Adherent	.469	.416	
Closed Question		.751	
Simple Reflection		.743	
Giving Information		.480	
Complex Reflection	.392	.467	
MI-Non-adherent		.433	
Open Question			.667
Eigenvalues	4.075	2.576	1.219
Percentage of total variance	34.044	21.470	10.155
Number of MITI measures	7	6	3

Appendix B: Figures:

Figure 1. *Percentage of the Sample that met MI Proficiency Standards.*



References

Acosta, M., Haller, D. L., & Ingersoll, A. K. (2010). The Science of Motivational Interviewing. *Handbook of Motivation and Change: A Practical Guide for Clinicians*, 237.

Aharonovich, E., Amrhein, P. C., Bisaga, A., Nunes, E. V., & Hasin, D. (2008). Cognition, Commitment Language, and Behavioral Change Among Cocaine Dependent Patients. *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *22*(4), 557–562. https://doi.org/10.1037/a0012971

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology*, *71*(5), 862–878. https://doi.org/10.1037/0022-006X.71.5.862

Armstrong, M. J., Mottershead, T. A., Ronksley, P. E., Sigal, R. J., Campbell, T. S., & Hemmelgarn, B. R. (2011). Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. *Obesity Reviews*, *12*(9), 709-723. doi: 10.1111/j.1467-789X.2011.00892.x

Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*, *73*(1), 99–106.

Baer, J. S., Beadnell, B., Garrett, S. B., Hartzler, B., Wells, E. A., & Peterson, P. L. (2008). Adolescent change language within a brief motivational intervention and substance use outcomes. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *22*(4), 570–575. https://doi.org/10.1037/a0013022

Beidas, R., Kendall, P.C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice.* 17(1):1–30. doi: 10.1111/j.1468-2850.2009.01187.x

Boardman, T., Catley, D., Grobe, J. E., Little, T. D., & Ahluwalia, J. S. (2006). Using motivational interviewing with smokers: do therapist behaviors relate to engagement and therapeutic alliance? *Journal of Substance Abuse Treatment*, *31*(4), 329–339. https://doi.org/10.1016/j.jsat.2006.05.006

Brown, J.D. (2000). What issues affect Likert-scale questionnaire formats? JALT Testing & Evaluation SIG, 4, 27-30.

Burnam, M.A., Hepner, K.A., & Miranda, J. (2009). Future research on psychotherapy practice in usual care. *Administration and Policy in Mental Health and Mental Health Services Research.* 36:1–5. doi: 10.1007/s10488-009-0254-7.

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *Journal of consulting and clinical psychology*, *71*(5), 843.

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT Alcohol Consumption Questions (AUDIT-C): An Effective Brief

Screening Test for Problem Drinking. *Archives of Internal Medicine*, *158*(16), 1789–1795. https://doi.org/10.1001/archinte.158.16.1789

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*

COMBINE Study Research Group. (2003). Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: rationale and methods. *Alcoholism, clinical and experimental research*, *27*(7), 1107.

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-Type Scales in the Social Sciences. *Journal of Adult Education*, *40*(1), 19–22.

D'Amico, E. J., Houck, J. M., Hunter, S. B., Miles, J. N., Osilla, K. C., & Ewing, B. A. (2015). Group motivational interviewing for adolescents: change talk and alcohol and marijuana outcomes. *Journal of consulting and clinical psychology*, *83*(1), 68.

Damschroder, L.J., Hagedorn, H.J. (2011). A guiding framework and approach for implementation research in substance use disorders treatment. *Psychology of Addictive Behaviors.* 25(2):194–205. doi: 10.1037/a0022284.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795–805. https://doi.org/10.1037/0021-9010.84.5.795

Fernandes, M., & Randall, D. (1991). The social desirability response bias in ethics research. Journal of Business Ethics, 10 (11), 805-807.

Forsberg, L., Källmén, H., Hermansson, U., Berman, A. H., & Helgason, Á. R. (2007). Coding Counsellor Behaviour in Motivational Interviewing Sessions: Inter-Rater

Reliability for the Swedish Motivational Interviewing Treatment Integrity Code (MITI). *Cognitive Behaviour Therapy*, *36*(3), 162-169.

Gentilello, L. M., Rivara, F. P., Donovan, D. M., Jurkovich, G. J., Daranciang, E., Dunn, C. W., … Ries, R. R. (1999). Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. *Annals of Surgery*, *230*(4), 473-480; discussion 480-483.

Glynn, L. H., & Moyers, T. B. (2010). Chasing change talk: the clinician's role in evoking client language about change. *Journal of Substance Abuse Treatment*, *39*(1), 65–70. https://doi.org/10.1016/j.jsat.2010.03.012

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23.

Handmaker, N. S., Miller, W. R., & Manicke, M. (1999). Findings of a pilot study of motivational interviewing with pregnant drinkers. *Journal of studies on alcohol*, *60*(2), 285-287.

Harding, R., Dockrell, M. J. D., Dockrell, J., & Corrigan, N. (2001). Motivational interviewing for HIV risk reduction among gay men in commercial and public sex settings. *AIDS care*, *13*(4), 493-501.

Head, M.L., Holman, L., Lanfear, R., Kahn, A.T., Jennions, M.D. (2015) The Extent and Consequences of P-Hacking in Science. PLoS Biology 13(3): e1002106. doi:10.1371/journal.pbio.1002106

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annual. Review of Clinical Psychology*, *1*, 91-111. doi: 10.1146/annurev.clinpsy.1.102803.143833

Hodgins, D. C., Ching, L. E., & McEwen, J. (2009). Strength of commitment language in motivational interviewing and gambling outcomes. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *23*(1), 122–130. https://doi.org/10.1037/a0013010

Houck, J.M., & Moyers, T.B. (2008). What you do matters: Therapist influence on client behavior during motivational interviewing sessions. Poster session presented at the International Addiction Summit, Melbourne, Australia.

Humeniuk, R., Ali, R., Babor, T.F., et al. (2008) Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST). Addiction. 103(6): 1039-1047.

Kadden, R., Carbonari, J., Litt, M., Tonigan, S., & Zweben, A. (1998). Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. *Alcohol Clinical Experimental Research*, *22*, 1300-1311.

Longabaugh, R., Woolard, R. E., Nirenberg, T. D., Minugh, A. P., Becker, B., Clifford, P. R., … Gogineni, A. (2001). Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. *Journal of Studies on Alcohol*, *62*(6), 806–816. https://doi.org/10.15288/jsa.2001.62.806

Lundahl, B., Moleni, T., Burke, B. L., Butters, R., Tollefson, D., Butler, C., & Rollnick, S. (2013). Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. *Patient education and counseling*, *93*(2), 157-168. doi:

Madson, M. B., & Campbell, T. C. (2006). Measures of fidelity in motivational enhancement: A systematic review. *Journal of substance abuse treatment*, *31*(1), 67-73. doi:10.1016/j.jsat.2006.03.010.

Madson, M. B., Campbell, T. C., Barrett, D. E., Brondino, M. J., & Melchert, T. P. (2005). Development of the motivational interviewing supervision and training scale. *Psychology of Addictive Behaviors*, *19*(3), 303.

McLeod, B. D., Southam-Gerow, M., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review*, *38*(4), 541.

McMurran, M. (2009). Motivational interviewing with offenders: A systematic review. *Legal and Criminological Psychology*, *14*(1), 83-100.

Messer, S. B., & Wampold, B. E. (2002). Let's Face Facts: Common Factors Are More Potent Than Specific Therapy Ingredients. *Clinical Psychology: Science and Practice*, *9*(1), 21–25. https://doi.org/10.1093/clipsy.9.1.21

Miller, W. R. (1995). *Motivational enhancement therapy manual: A clinical research guide for therapists treating individuals with alcohol abuse and dependence* (Vol. 2). DIANE Publishing.

Miller, W. R. (2001). When is it motivational interviewing? Addiction, 96, 1770– 1771.

Miller, W. R., & Moyers, T. B. (2015). The forest and the trees: relational and specific factors in addiction treatment. *Addiction*, *110*(3), 401-413.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist*, *64*(6), 527.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

Miller, W. R., & Sanchez, V. C. (1994). Motivating young adults for treatment and lifestyle change. University of Notre Dame Press.

Miller, W. R., Taylor, C. A., & West, J. c. (1980). Focused versus broad-spectrum behavior therapy for problem drinkers. *Journal of Consulting and Clinical Psychology*, *48*(5), 590–601.

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*(6), 1050–1062. https://doi.org/10.1037/0022-006X.72.6.1050

Monti, P. M., & Rohsenow, D. J. (1999). Coping-skills training and cue-exposure therapy in the treatment of alcoholism. *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, *23*(2), 107–115.

Moyers, T. B., Houck, J., Rice, S. L., Longabaugh, R., & Miller, W. R. (2016). Therapist empathy, combined behavioral intervention, and alcohol outcomes in the COMBINE research project. *Journal of Consulting and Clinical Psychology*, *84*(3), 221–229. https://doi.org/10.1037/ccp0000074

Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of substance abuse treatment*, *28*(1), 19-26.

Moyers, T., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy*, *31*(02), 177-184.

Moyers, T. B., Rowell, L. N., Manuel, J. K., Ernst, D., & Houck, J. M. (2016). The Motivational Interviewing Treatment Integrity Code (MITI 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*.

Moyers, T. B., Manuel, J. K., Wilson, P. G., Hendrickson, S. M. L., Talcott, W., & Durand, P. (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, *36*(2), 149–162. https://doi.org/10.1017/S1352465807004055

Parsons, J. T., Rosof, E., Punzalan, J. C., & Maria, L. D. (2005). Integration of motivational interviewing and cognitive behavioral therapy to improve HIV medication adherence and reduce substance use among HIV-positive men and women: results of a pilot project. *AIDS Patient Care & STDs*, *19*(1), 31-39.

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice*, *18*(2), 148-153. doi:10.1111/j.1468-2850.2011.01246.x.

Perepletchikova, F., Kazdin, A.E. (2005) Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice.* 12:365–383. doi: 10.1093/clipsy/bpi045.

Prochaska, J. O., & Norcross, J. C. (2007). Systems of psychotherapy: A transtheoretical approach. South Melbourne, Australia: Thompson Brooks/Cole.

Rogers, C. R. (1951). *Client-centered Therapy: Its Current Practice, Implications, and Theory, with Chapters*. Houghton Mifflin.

Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, *23*(04), 325-334.

Rubak, S., Sandbæk, A., Lauritzen, T., & Christensen, B. (2005). Motivational interviewing: a systematic review and meta-analysis. *British Journal of General Practice*, *55*(513), 305-312. doi:

Skinner, H. A. (1982). The drug abuse screening test. *Addictive Behaviors*, *7*(4), 363–371.

Sobell, L.C., Sobell, M.B. (1992). Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten RA, Allen JP, eds. Measuring Alcohol Consumption: Psychosocial and Biological Methods. Totowa, NJ: Humana Press.

Southam-Gerow, M. A., & McLeod, B. D. (2013). Advances in applying treatment integrity research for dissemination and implementation science: Introduction to special issue. *Clinical Psychology: Science and Practice*, *20*(1), 1-13. doi: 10.1111/j.1468-2850.2006.00001.x.

Tollison, S. J., Lee, C. M., Neighbors, C., Neil, T. A., Olson, N. D., & Larimer, M. E. (2008). Questions and reflections: the use of motivational interviewing microskills in a peer-led brief alcohol intervention for college students. *Behavior Therapy*, *39*(2), 183–194. https://doi.org/10.1016/j.beth.2007.07.001

Vader, A. M., Walters, S. T., Prabhu, G. C., Houck, J. M., & Field, C. A. (2010). The language of motivational interviewing and feedback: counselor language, client language, and client drinking outcomes. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *24*(2), 190–197. https://doi.org/10.1037/a0018749

VanBuskirk, K. A., & Wetherell, J. L. (2014). Motivational interviewing with primary care populations: a systematic review and meta-analysis. *Journal of behavioral medicine*, *37*(4), 768-780.

Vasilaki, E. I., Hosier, S. G., & Cox, W. M. (2006). The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. *Alcohol and Alcoholism*, *41*(3), 328-335.

Webb, C.A., DeRubeis, R.J., Barber, J.P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology.* 78(2):200–211. doi: 10.1037/a0018912.

Witkiewitz, K., Falk, D. E., Kranzler, H. R., Litten, R. Z., Hallgren, K. A., O'malley, S. S., & Anton, R. F. (2014). Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials. *Alcoholism: clinical and experimental research*, *38*(11), 2826-2834.