Published in final edited form as:
Psychol Addict Behav. 2012 December ; 26(4): 994–1000. doi:10.1037/a0027987.

Assessing Motivational Interviewing Integrity for Group Interventions with Adolescents

Elizabeth J. D'Amico, **Karen C. Osilla**, **Jeremy N.V. Miles**, **Brett Ewing**, **Kristen Sullivan**, **Kristin Katz**, and **Sarah B. Hunter**

Abstract

The group format is commonly used in alcohol and other drug (AOD) adolescent treatment settings, but little research exists on the use of Motivational Interviewing (MI) in groups. Further, little work has assessed the integrity of MI delivered in group settings. This study describes an approach to evaluate MI integrity using data from a group MI intervention for at-risk youth. Using the Motivational Interviewing Treatment Integrity (MITI) scale 3.1, we coded 140 group sessions, led by three different facilitators. Four trained coders assessed the group sessions. Agreement between raters was evaluated using a method based on limits of agreement, and key decisions used to monitor and calculate group MI integrity are discussed. Results indicated that there was adequate agreement between raters; we also found differences on use of MI between the MI-intervention group and a usual care group on MI global ratings and behavioral counts. This study demonstrates that it is possible to determine whether group MI is implemented with integrity in the group setting and that MI in this setting is different from what takes place in usual care.

The group format is commonly used to treat adolescents who use alcohol and other drugs (AOD) (Kaminer, 2005; Vaughn & Howard, 2004) even though little is known about the distinguishing factors of effective and ineffective group interventions (Engle, Macgowan, Wagner, & Amrhein, 2010). Recent work with at-risk adolescents indicates that motivational interviewing (MI) interventions (Miller & Rollnick, 2002; Rollnick, Miller, & Butler, 2008) can be successful as MI offers a collaborative, non-judgmental, and non-confrontational approach (Naar-King & Suarez, 2010). Although studies of youth receiving individual MI have shown effectiveness (Baer, Garrett, Beadnell, Wells, & Peterson, 2007; D'Amico, Miles, Stern, & Meredith, 2008; Monti, et al., 2007; Spirito, et al., 2004), there are only four published group MI studies (Bailey, Baker, Webster, & Lewin, 2004; D'Amico, Osilla, & Hunter, 2010; Engle, et al., 2010; Schmiege, Broaddus, Levin, & Bryan, 2009). Furthermore, MI treatment integrity has been mainly conducted with individual interventions (Baer, et al., 2004; Martino, Ball, Nich, Frankforter, & Carroll, 2008; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005; Vader, Walters, Prabhu, Houck, & Field, 2010). Only one study to date has evaluated MI integrity with adolescent group sessions (Engle, et al., 2010) using measures of MI competence and adherence, such as the Motivational Interviewing Treatment Integrity scale (Moyers, Martin, Manuel, Miller, & Ernst, 2010).

Treatment integrity is important to ensure that a therapy is delivered as intended by the treatment developers (Beidas & Kendall, 2010; Godley, Garner, Smith, Meyers, & Godley, 2011; Rakovshik & McManus, 2010). The Motivational Interviewing Treatment Integrity (MITI) scale (Moyers, et al., 2010) is a widely used measure of MI competence and adherence. Engle et al. (2010) used the MITI to examine the group process and found that

Author Correspondence: Elizabeth J. D'Amico, Ph.D. RAND Corporation 1776 Main St., PO Box 2138 Santa Monica, CA 90407-2138 Phone: 310-393-0411 x6487 Fax: 310-260-8150 elizabeth_d'amico@rand.org.

group facilitator empathy was associated with more positive commitment language (e.g., *"I'm quitting for the summer")*, which was then associated with reduced AOD use (Engle, et al., 2010). Although this study was pioneering, further studies are needed as only 19 sessions were coded, only one of the MITI global ratings (empathy) was assessed, and there was no comparison group. Thus, to move the field forward, more detailed treatment integrity analyses (i.e., more group sessions and a comparison group) are needed so that we can assess how MI groups differ from other adolescent groups. .

Finally, in order for treatment integrity to be successfully measured, alternative measurements to intra-class correlation (ICC) are needed to assess inter-rater agreement (e.g., Shrout & Fleiss, 1979). Calculating the ICC is not always feasible as there may be different sets of raters for each session or there may be missing data (e.g., one session might be coded by all raters, but another session may not). Thus, other approaches that compensate for these limitations must be utilized in order for studies to quantify inter-rater agreement.

The current study addresses these gaps by using the global and behavioral counts on the MITI to 1) assess agreement between raters, 2) calculate inter-rater agreement using two alternative methods to the ICC, and 3) evaluate the use of MI between a group led by facilitators trained and supervised in MI as compared to a group with a self-help and Alcoholics Anonymous focus. This is an important first step in understanding whether the MITI is feasible for use in group settings and determining MI integrity in this setting. We hypothesized that coders would reliably assess MI behaviors and that trained and supervised MI facilitators would deliver more MI consistent behavior than the facilitator in the usual care (non-MI) group sessions.

Method

Study Overview

This study was conducted as part of a randomized clinical trial (D'Amico, Hunter, et al., 2010). Procedures were approved by the research institution's Internal Review Board. We collaborated with the Council on Alcoholism and Drug Abuse, a nonprofit organization in Santa Barbara, California that operates a Teen Court (TC) for first-time offending youth. Based on earlier assessments, adolescents who do not need intensive treatment can participate in TC in lieu of formal juvenile justice processing. As part of TC, AOD offending youth are required to attend six AOD awareness groups, along with other sanctions (e.g., community service). Adolescents who successfully complete their sentence have the offense expunged from their record.

The AOD awareness groups occur weekly. Entrance to the groups is based on rolling admission as each session can stand alone without a teen having to complete a previous session—that is attending session 2 does not require information from session 1. Teens enter the sessions based on their time of referral to the program. Teens were randomized to either a usual care (UC) control condition or the experimental MI group intervention called Free Talk (FT). Teens were automatically assigned to attend the UC groups if they or their parent refused participation. Refusals (10%) were mostly due to lacking time or transportation to complete a baseline survey before their first group session. There were no demographic or offense differences between refusers and participants. See Table 1 for FT participant characteristics.

The current study

This study focuses on the facilitator's behavior and whether this behavior can be reliably coded during adolescent group sessions. We provide detailed information on monitoring of a group MI intervention with at-risk adolescents (n=102) across a large number of group

sessions (n=140) and compare these MI adolescent groups to usual care groups using the MITI.

Intervention Condition: Free Talk Groups

Free Talk was developed as part of a Stage 1 study (Rounsaville, Carroll, & Onken, 2001) where each group session was iteratively tested to determine feasibility and acceptability of intervention content (D'Amico, Osilla, et al., 2010). The facilitator manual for FT included a protocol for each session and utilized a MI approach. At the beginning of each session, the facilitator discussed the group guidelines (e.g., confidentiality, respect for others in the group). These were provided in a MI consistent way (e.g., ensuring members were respectful of one another). FT is a manualized intervention that provides some pyschoeducation and focuses on encouraging change talk (see D'Amico, Osilla et al., 2010 for content). For example, throughout all sessions, the focus is on providing reflections, asking open ended questions to increase collaboration, affirming and supporting the adolescents to increase support and autonomy, and increasing change talk in the group by being aware of the DARN-C (Desire, Ability, Reason, Need and Commitment; (Amrhein, 2009). We also used tools to promote behavioral change such as the decisional balance activity and willingness and confidence rulers (Ingersoll, Wagner, & Gharib, 2006; Miller & Rollnick, 2002). Sessions lasted 55 minutes and average group size was 4 adolescents (mean=4.48; SD=1.98).

Free Talk training and integrity monitoring

The FT sessions were led by one of three facilitators (all female and white) who were psychology doctoral students with prior at-risk teen work experience. They received 40 hours of MI and FT training delivered by ED and KO, clinical psychologists affiliated with the Motivational Interviewing Network of Trainers (MINT).

Facilitators were instructed to use MI to best fit a group format while still attending to individuals (D'Amico, Feldstein Ewing, et al., 2010; Velasquez, Stephens, & Ingersoll, 2006). For example, reflections were often stated to address the group process (e.g., "Many of you have been making positive changes in your lives"), and yet facilitators also responded to the varying individual experiences and needs (e.g., rolling with the resistance of one youth, while trying to actively maintain the commitment language of another). In addition, facilitators monitored participant feedback and redirected negative or unhelpful comments to create a safe place for participation and mutual respect.

All FT groups were digitally audio recorded. MINT trainers reviewed recordings and provided one-hour weekly supervision to facilitators. The MITI was used to monitor performance (i.e., provide feedback during supervision).

Control Condition: Usual Care Groups

The usual care (UC) condition was led by one facilitator who was male and Hispanic. The curriculum followed an abstinence-based Alcoholics Anonymous approach. Topics included group check-in/discussion of personal triggers, consequences of AOD use, educational videos, discussion of personal experiences with AOD use, and myths about AOD use. Like the FT groups, session and each session lasted about 55 minutes.

Coding

Overview—One hundred and forty sessions (70 FT and 70 UC sessions) were coded by four psychology doctoral students. Per the MITI protocol, a randomly selected 20-minute segment was coded for both FT and UC sessions (Moyers, et al., 2010). FT sessions were

coded using digital audio recordings and UC sessions were coded from live observation because not all teens attending these groups were study participants so it was not possible to record these sessions. Raters were trained to not disrupt the UC groups by arriving early, sitting off to the side, not interacting with teens, and being discreet with coding materials.

MITI—The MITI is a widely used instrument for coding competency and adherence to MI and has been used in numerous studies to assess MI integrity (Jensen, et al., 2011; Moyers, et al., 2005; Tollison, et al., 2008; Turrisi, et al., 2009). It was specifically designed for clinical trials and is a reliable and valid tool (Madson & Campbell, 2006; Moyers, et al., 2005). Integrity is measured through the use of global scores (e.g., amount of collaboration) and behavioral counts (e.g., number of reflections) (Moyers, et al., 2010). The MITI is used to assess the therapist's behavior during a session and focuses solely on the therapist's behavior; thus the MITI can easily be utilized to assess a group facilitator's behavior (e.g., Engle, et al., 2010). MITI 3.1 has five global scales (collaboration, empathy, evocation, autonomy/support, and direction) that are scored on a scale from 1 (low) to 5 (high). MITI competency is defined as an average of 4 on the global ratings (Moyers, et al., 2010). As noted in the MITI manual, collaboration occurs when there is little power differential, there is agreement on goals, and clients are encouraged to share the talking. Empathy occurs when the facilitator expresses client understanding and attempts to understand client point of view. Evocation occurs when the facilitator encourages clients to brainstorm reasons and ideas for how to change. Autonomy/support occurs when the facilitator emphasizes and supports client's personal choice. Direction occurs when the facilitator exerts influence on the session and generally does not miss opportunities to direct client toward the target behavior or referral question (Moyers, et al., 2010). In the group setting, the facilitator responds to group members, thus group members influence the facilitator's behavior. For example, if a facilitator is not collaborative and does not ask open ended questions, then there will be less participation and less sharing of the talking. Facilitator responsiveness to group member behavior is captured by the MITI global scores.

The rater also counts specific behaviors that occur during each coded segment including open-ended questions, closed-ended questions, MI adherent (e.g., "if it's ok with you, I'd like to share some information with you") and non-adherent statements (e.g., "*you need to stop drinking"*), and simple (e.g., "*some of you are ready to make changes"*) and complex reflections (e.g., "some of you are hoping that by making changes, things will improve in your lives"). Whereas global scores have a range limit (1-5), behavioral counts have no upper end on the scale; thus these scores can vary by session to a greater degree.

Coding training

Four raters received about 40 hours of training, which included a half-day MITI training and practice coding assignments (http://casaa.unm.edu/codinginst.html). Similar to other studies (Moyers, et al., 2005; Tollison, et al., 2008), raters met weekly to discuss discrepancies. Inter-rater agreement was stable over time. Raters coded both UC and FT group sessions.

Inter-rater agreement for MITI

All 140 sessions were coded by at least one rater, with 48 (34%) sessions coded by two raters (19 UC and 29 FT) and 25 (18%) FT sessions coded by three raters. We did not have three raters code UC sessions given that these groups were coded "live" and this would have been disruptive. Raters were not the same two people each time as they were assigned to code sessions randomly. Thus, the ICC could not be calculated as different raters coded different sessions and ICC is not defined in the presence of missing data. In addition, using the ICC in this case would not be an accurate comparison as it is like reporting correlations among items, rather than a coefficient alpha. Interpretation of these coefficients would also

be difficult because many of them have small sample sizes and hence we do not have a sense of the sampling distribution of the statistic.

We therefore used two different methods to quantify agreement between raters. Our first method was distribution free; we calculated the difference between each rater's rating of a session and the mean rating for that session. To see how close the majority of ratings were to the mean, we ordered these differences from the smallest to the largest. We then took the 95th percentile (e.g., if there were 100 differences, we took the 95th rating, referred as D95 in Table 2). Thus, the D95 value indicates that 95% of differences between each rater and the mean rating are smaller than this value. The D95 is more appropriate for dichotomous variables as it makes no distributional assumptions.

In our second method, we used a normal distribution approach of calculating the within-session standard deviation (WSSD) to provide an estimate of the difference between raters (Altman & Bland, 1983; Bland & Altman, 1986, 2007). The WSSD is closely related to the average of the difference between raters $-\Sigma\left(x-\bar{x}\right)/N$ (where $\bar{x}$ is the mean for the session, and x represents the set of ratings). The WSSD is given by the formula for the standard deviation $\left(\Sigma\left(x-\bar{x}\right)^2/(N-1)\right)^{0.5}$. Because the WSSD squares the differences between each rating and the mean, larger deviations between raters lead to proportionally larger estimates. If we assume differences are normally distributed, we expect 95% of ratings to lie within about two standard deviations of that mean. For interpretation purposes, we therefore report 2*WSSD, which allows us to compare the two methods.

These two methods make different assumptions about the distribution of ratings. The D95 method does not make distributional assumptions, but it does not use all of the information that is available regarding the values. In contrast, the WSSD method uses all the information and assumes that data are normally distributed and measured on a continuous scale. Using two methods allowed us to examine inter-rater agreement sensitivity and compare the results of the two methods.

These agreement measures have two advantages over the ICC. First, the ICC is a ratio of between group variance to total variance. The ICC depends not just on the extent of the agreement between raters, but also on the variance of the measures. Thus, the ICC could be lower or higher depending on the level of variance even if the rater agreement level was the same in both cases. Second, the ICC presents agreement on a scale from 0.0 (no agreement) to 1.0 (perfect agreement). Although the scale allows for a comparison across studies, the ICC has little interpretive value of how actual scores differ between raters because the units of the ICC are not the same as the units on the scale. This variation is very important for scales like the MITI that have items ranging from 1 to 5 points because it provides data on how to recalibrate coding to ensure that they match.

Results

Inter-rater agreement

We calculated the limits of agreement between the raters for each of the MITI dimensions using the D95 and 2*WSSD methods. Table 2 shows ratings for both groups combined across all four raters. For MITI global ratings using the D95 method, raters were within 0.5 points of the mean for the session 95% of the time. Using the 2*WSSD method, the level of agreement was similar, with expected 95% limits ranging from 0.50 (Autonomy/Support and Empathy) to 0.62 (Evocation).

For the behavioral counts, there was no upper limit; thus, we expected to see larger differences. Both approaches yielded comparable limits of agreement. The D95 limits ranged from 3.56 (Complex Reflections) to 5.56 (MI Adherent and Closed Questions). The 2*WSSD limits ranged from 3.26 (Complex Reflections) to 5.54 (Closed Questions). As one would expect, MI Non-Adherent counts had a lower D95 (0.50) and 2*WSSD (0.36). There are no published MITI bench marks for what consists of good agreement. For this study, we consider within a point on the MITI global scores and within 6 points on the behavioral counts as good levels of agreement.

MI differences by group

Table 3 shows differences between the means of all raters across groups. Overall, both global and behavioral count scores were higher for the FT group compared to the UC group, indicating greater MI integrity in the FT group sessions.

Discussion

This study is one of the first to assess MI integrity using data from an adolescent group intervention. First, this study found that group MI can be efficiently and reliably assessed using alternatives to ICC. We used two innovative methods to address agreement and both found comparable results. These methods are more useful than the ICC because they can be calculated when using multiple coders and are better able to quantify the difference between multiple raters and provide us with a more direct interpretation of the level of agreement. For example, a D95 or 2*WSSD of 2 tells us that 95% of raters are expected to be within 2 points of the mean of all raters. A clinical judgment must then be made to determine whether the raters are sufficiently close to agreement, which depends upon on the units of the original measure. A difference of 2 points on global ratings is very large compared to behavioral counts. Thus, our findings of less than .62 difference between a rater and the mean score on the global ratings and 5.56 on the behavioral counts suggest that the raters were in fairly close agreement. Using these indices of interrater agreement could be useful in clinical practice as it bridges the gap between research and practice by allowing for real world situations, such as multiple raters or supervisors.

We found that it is feasible to train facilitators to conduct MI in an adolescent group setting and that MI can be measured and implemented with integrity. Overall, global scores for the FT group were in the competent range and behavioral counts for MI behaviors were high. We expected this given the intensive training and supervision; MI fidelity might not be as high with less supervision. Also, future research could measure integrity across more facilitators.

We found large differences between the FT and UC groups on collaboration, empathy, evocation and autonomy/support suggesting that FT groups were more likely to encourage power sharing during the session, had repeated efforts to gain understanding of teens' viewpoints, had more acceptance of teens' reasons for change, and were more likely to emphasize support of client autonomy. The behavioral count data showed a similar pattern with more simple and complex reflections and open-ended questions in FT groups. More MI non-adherent behavior was observed in the UC groups, such as confronting and advising without permission compared to the FT groups. Findings are not surprising given that the UC group facilitator did not receive MI training nor supervision but rather followed an Alcoholics Anonymous group treatment approach. Our study findings suggest that, similar to MI research with individuals (Moyers, et al., 2005), group MI could make the overall group process more collaborative and therefore more effective and may reduce the likelihood of iatrogenic effects that are sometimes seen in groups of at-risk youth (e.g., Dishion, McCord, & Poulin, 1999; Dodge, Dishion, & Lansford, 2006). For example, as

shown with individual work (Magill, Apodaca, Barnett, & Monti, 2010; Moyers, Martin, Houck, Christopher, & Tonigan, 2009), the increased evocation and autonomy/support may provide adolescents with a safe place to explore reasons for change without fear of being forced change; the guiding and empathic style of MI may then help move participants toward positive behavior change.

This study had limitations. First, raters were not blind to study condition. Similar to other MI brief interventions (e.g., Hettema, Steele, & Miller, 2005), FT sessions followed a protocol and could be easily distinguished from other types of sessions. Second, our UC control condition had to be coded by live observation; thus, it is possible that some information was missed. However, raters coded both UC and FT with one pass so we feel confident that raters captured most of the behaviors during the UC sessions. Moreover, our results indicated that raters were consistent in their ratings across both groups and had high agreement, suggesting that they were not biased. Third, no prior research has employed these statistical methods to address inter-rater agreement for the MITI; thus no consensus exists regarding how much difference among raters might be "acceptable". We do not consider this a problem as cutoff values may vary across studies and may not be quantifiable, and the use of such values could limit researchers from interpreting results objectively. Thus, it is important to use clinical judgment when evaluating the level of agreement with these methods. Finally, we did not assess group cohesion or engagement; more work is needed in this area.

In sum, this study takes an important first step in documenting group MI with at-risk youth. This study is among the first to provide within session standard deviations for all global and behavioral counts that comprise the MITI scoring system. Thus, this work can provide a potential resource to practitioners and researchers who do group work and are interested in using MI and measuring MI integrity in this setting. The next step in this work is to examine intervention efficacy; we are currently conducting a randomized trial to assess the short-term effects of this intervention on AOD use among at-risk teens.

Acknowledgments

We would like to thank the Council on Alcoholism and Drug Abuse for their support of this project. We would also like to thank Cally Sprague, Susana Lopez for their help on the study. The current study was funded by a grant from the National Institute of Drug Abuse (R01DA019938) to Elizabeth D'Amico.

References

Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. The Statistician. 1983; 32:307–317.

Amrhein, P,C. A training manual for coding client commitment language. Columbia University; New York: 2009.

Baer JS, Garrett SB, Beadnell B, Wells EA, Peterson PL. Brief motivational intervention with homeless adolescents: Evaluating effects on substance use and service utilization. Psychology of Addictive Behaviors. 2007; 21:582–586. [PubMed: 18072842]

Baer JS, Roesengren DB, Dunn CW, Wells EA, Ogle RL, Hartzler B. An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. Drug & Alcohol Dependence. 2004; 73:99–106. [PubMed: 14687964]

Bailey KA, Baker AL, Webster RA, Lewin TJ. Pilot randomized controlled trial of a brief alcohol intervention group for adolescents. Drug and Alcohol Review. 2004; 23(2):157–166. [PubMed: 15370021]

Beidas RS, Kendall PC. Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. Clinical Psychology: Science and Practice. 2010; 17:1–30. [PubMed: 20877441]

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986:307–310.

Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. Journal of Biopharmaceutical Statistics. 2007; 17:571–582. [PubMed: 17613642]

D'Amico, EJ.; Feldstein Ewing, SW.; Engle, B.; Hunter, SB.; Osilla, KC.; Bryan, A. Group alcohol and drug treatment. In: Naar-King, S.; Suarez, M., editors. Motivational interviewing with adolescents and young adults. The Guilford Press; New York, NY: 2010. p. 151-157.

D'Amico, EJ.; Hunter, SB.; Osilla, KC.; Miles, JNV.; Munjas, B.; Saunders, J. Group MI for first time alcohol and marijuana offending adolescents; Paper presented at the Research Society on Alcoholism; San Antonio, TX. 2010.

D'Amico EJ, Miles JNV, Stern SA, Meredith LS. Brief motivational interviewing for teens at risk of substance use consequences: A randomized pilot study in a primary care clinic. Journal of Substance Abuse Treatment. 2008; 35:53–61. [PubMed: 18037603]

D'Amico EJ, Osilla KC, Hunter SB. Developing a group motivational interviewing intervention for adolescents at-risk for developing an alcohol or drug use disorder. Alcoholism Treatment Quarterly. 2010; 28:417–436. [PubMed: 21113392]

Dishion TJ, McCord J, Poulin F. When interventions harm: Peer groups and problem behavior. American Psychologist. 1999; 54:755–764. [PubMed: 10510665]

Dodge, K.; Dishion, TJ.; Lansford, JE. Deviant peer influences in programs for youth. The Guilford Press; New York, NY: 2006.

Engle B, Macgowan MJ, Wagner EF, Amrhein P. Markers of marijuana use outcomes within adolescent substance abuse group treatment. Research on Social Work Practice. 2010; 20:271–282.

Godley SH, Garner BR, Smith JE, Meyers RJ, Godley MD. A large-scale dissemination and implementation model for evidence-based treatment and continuing care. Clinical Psychology: Science and Practice. 2011; 18:67–83. [PubMed: 21547241]

Hettema J, Steele J, Miller WR. Motivational interviewing. Annual Review of Clinical Psychology. 2005; 1(1):91–111.

Ingersoll, KS.; Wagner, CC.; Gharib, S. Motivational groups for community substance abuse programs (3rd edition). Substance Abuse Mental Health Services Administration; Rockville, MD: 2006.

Jensen CD, Cushing CC, Aylward BS, Craig JT, Sorell DM, Steele RG. Effectiveness of motivational interviewing interventions for adolescent substance use behavior change: A meta-analytic review. Journal of Consulting and Clinical Psychology. Jul 4.2011 Online First Publication. 2011.

Kaminer Y. Challenges and opportunities of group therapy for adolescent substance abuse: A critical review. Addictive Behaviors. 2005; 30(9):1765–1774. [PubMed: 16051443]

Madson MB, Campbell TC. Measures of fidelity in motivational enhancement: A systematic review. Journal of Substance Abuse Treatment. 2006; 31(1):67–73. [PubMed: 16814012]

Magill M, Apodaca TR, Barnett NP, Monti PM. The route to change: Within-session predictors of change plan completion in a motivational interview. Journal of Substance Abuse Treatment. 2010; 38:299–305. [PubMed: 20149571]

Martino S, Ball SA, Nich C, Frankforter TL, Carroll KM. Community program therapist adherence and competence in motivational enhancement therapy. Drug and Alcohol Dependence. 2008; 96(1):37–48. [PubMed: 18328638]

Miller, WR.; Rollnick, S. Motivational interviewing: Preparing people for change. 2nd ed. Guilford Press; New York, NY: 2002.

Monti PM, Barnett NP, Colby SM, Gwaltney CJ, Spirito A, Rohsenow DJ, et al. Motivational interviewing versus feedback only in emergency care for young adult problem drinking. Addiction. 2007; 102(8):1234–1243. [PubMed: 17565560]

Moyers TB, Martin T, Houck JM, Christopher PJ, Tonigan JS. From in-session behaviors to drinking outcomes: A causal chain for motivational interviewing. Journal of Consulting & Clinical Psychology. 2009; 77:1113–1124. [PubMed: 19968387]

Moyers TB, Martin T, Manuel JK, Hendrickson SML, Miller WR. Assessing competence in the use of motivational interviewing. Journal of Substance Abuse Treatment. 2005; 28:19–26. [PubMed: 15723728]

Moyers, TB.; Martin, T.; Manuel, JK.; Miller, WR.; Ernst, D. Revised Global Scales:Motivational Interviewing Treatment Integrity 3.1.1 (MITI 3.1.1). 2010. Retrieved from http://casaa.unm.edu/mimanuals.html

Naar-King, S.; Suarez, M. Motivational interviewing with adolescents and young adults. The Guilford Press; New York, NY: 2010.

Rakovshik SG, McManus F. Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. Clinical Psychology Review. 2010; 30:496–516. [PubMed: 20488599]

Rollnick, S.; Miller, WR.; Butler, CC. Motivational interviewing in health care: Helping patients change behavior. Guilford Press; New York: 2008.

Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: Getting started and moving on from stage I. Clinical Psychology: Science and Practice. 2001; 8:133–142.

Schmiege SJ, Broaddus MR, Levin M, Bryan AD. Randomized trial of group interventions to reduce HIV/STD risk and change theoretical mediators among detained adolescents. Journal of Consulting and Clinical Psychology. 2009; 77(1):38–50. [PubMed: 19170452]

Shrout P, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428. [PubMed: 18839484]

Spirito A, Monti PM, Barnett NP, Colby SM, Sindelar H, Rohsenow D, et al. A randomized clinical trial of a brief motivational intervention for alcohol-positive adolescents treated in an emergency department. Journal of Pediatrics. 2004; 145:396–402. [PubMed: 15343198]

Tollison S, Lee C, Neighbors C, Neil T, Olson N, Larimer M. Questions and reflections: The use of motivational interviewing microskills in a peer-led brief alcohol intervention for college students. Behavior Therapy. 2008; 2:183–194. [PubMed: 18502251]

Turrisi R, Larimer M, Mallett K, Kilmer J, Ray A, Mastroleo N, et al. A randomized clinical trial evaluating a combined alcohol intervention for high-risk college students. Journal of Studies on Alcohol and Drugs. 2009; 70:555–567. [PubMed: 19515296]

Vader AM, Walters ST, Prabhu GC, Houck JM, Field CA. The language of motivational interviewing and feedback: Counselor language, client language, and client drinking outcomes. Psychology of Addictive Behaviors. 2010; 24:190–197. [PubMed: 20565145]

Vaughn MG, Howard MO. Adolescent substance abuse treatment: A synthesis of controlled evaluations. Research on Social Work Practice. 2004; 14(5):325–335.

Velasquez MM, Stephens NS, Ingersoll KS. Motivational interviewing in groups. Journal of Groups in Addiction and Recovery. 2006; 1:27–50.

Table 1

Characteristics of the adolescents in the Free Talk and Usual Care groups (n=102)

Variable	Overall
Mean Age (sd)	16.61 (1.02)
Gender (%)	
Male	69.61
Female	30.39
Race (%)	
Hispanic	43.14
Mixed	2.94
White	50.00
Other	3.92
Offense (%)	
Marijuana	42.16
Alcohol	52.94
Both	3.92
Other	0.98
Drug Used Most (%)	
Alcohol	34.31
Marijuana	58.82
Other	6.86
Past 30 Day Use (%)	
Heavy drinking	38.24
Marijuana use	63.73
More than 1 drug at a time	51.96
Significant Adult Use (%)	
Alcohol (4-7 days/week)	13.86
Marijuana (any)	13.13
Sexual Behavior (%)	
Ever had sex	65.69
Drug or alcohol use before sex	28.79
Unprotected sex	21.21
Delinquency in Past Year (%)	
Skipped school (in past 30 days)	42.16
Suspended or expelled from school	27.45
Carried a hidden weapon	9.80
Involved in fights	34.31

Table 2

MITI measures of agreement between raters and mean scores by rater across both the Free Talk and usual care groups

Number of Coded Sessions						
			Rater 1	**Rater 2**	**Rater 3**	**Rater 4**
			n=67	**n=45**	**n=73**	**n=53**
MITI Global Ratings						
	D95[a]	2*WSSD[b]	Mean	Mean	Mean	Mean
Evocation	0.50	0.62	3.5	3.5	3.6	3.5
Collaboration	0.50	0.60	4.2	4.0	3.6	3.7
Autonomy/Support	0.50	0.50	4.0	4.0	3.8	3.3
Direction	0.50	0.60	4.1	4.4	4.6	4.0
Empathy	0.50	0.50	4.0	4.0	3.9	3.3
MITI Behavioral Counts						
	D95[a]	2*WSSD[b]	Mean	Mean	Mean	Mean
Giving Information	4.50	4.16	11.0	9.2	12.1	9.0
MI Adherent	5.56	5.10	17.3	18.8	13.6	13.9
MI Non-Adherent	0.50	0.36	0.3	1.0	1.2	0.9
Closed Questions	5.56	5.54	15.5	15.8	19.4	13.5
Open Questions	4.44	4.92	21.4	19.5	17.4	16.9
Simple Reflections	5.00	4.62	10.6	12.0	11.6	11.1
Complex Reflections	3.56	3.26	8.3	8.7	7.0	6.4

[a]Note: D95 is the difference that is equal to or larger than 95% of all the differences

[b]2*WSSD is the mean within session standard deviation multiplied by 2.

Psychol Addict Behav. Author manuscript; available in PMC 2013 January 09.

Table 3

MITI means and standard deviations across raters by group

	Usual Care		Free Talk		
	Mean	**SD**	**Mean**	**SD**	**P-Value**
Global					
Evocation	2.2	0.9	4.4	0.4	<.001
Collaboration	2.4	0.8	4.7	0.5	<.001
Autonomy/Support	2.3	0.8	4.6	0.5	<.001
Direction	3.8	1.0	4.7	0.4	<.001
Empathy	3.0	1.1	4.4	0.5	<.001
Behavioral					
Giving Information	10.0	8.2	10.5	5.3	0.637
MI Adherent	7.3	5.2	20.2	6.6	<.001
MI Non-adherent	2.1	3.0	0.1	0.2	<.001
Closed Questions	17.6	12.4	15.4	6.4	0.186
Open Questions	8.3	6.6	25.0	8.4	<.001
Simple Reflections	6.7	6.6	14.0	5.4	<.001
Complex Reflections	1.0	6.6	11.3	4.4	<.001