

UC San Diego

UC San Diego Previously Published Works

Title

Feasibility and Reliability of a Coding System to Capture In-Session Group Behavior in Adolescents

Permalink

<https://escholarship.org/uc/item/176306hx>

Journal

Prevention Science, 17(1)

ISSN

1389-4986

Authors

Ladd, Benjamin O  
Tomlinson, Kristin  
Myers, Mark G  
et al.

Publication Date

2016

DOI

10.1007/s11121-015-0588-1

Peer reviewed

**Feasibility and Reliability of a Coding System to Capture In-session Group Behavior in  
Adolescents**

Benjamin O. Ladd<sup>1</sup>, Kristin Tomlinson<sup>2</sup>, Mark G. Myers<sup>2,3</sup> & Kristen G. Anderson<sup>1</sup>

<sup>1</sup>Adolescent Health Research Program, Department of Psychology, Reed College, Portland OR

<sup>2</sup>University of California San Diego, San Diego CA

<sup>3</sup>VA San Diego Healthcare System

Corresponding author: Benjamin Ladd

Mailing Address: Department of Psychology

3203 SE Woodstock Blvd

Portland, OR 97202

Email Address: [benladd@reed.edu](mailto:benladd@reed.edu)

Phone: (503) 517-7420

### Abstract

Limited research has explored the role of in-session behavior during motivational enhancement (ME) in group formats. The current study presents initial feasibility of assessing behavior of high school students ( $N = 425$ ) attending Project Options, a voluntary secondary drug and alcohol prevention program utilizing ME techniques. Building on previous research exploring client language supporting/opposing health behavior, student group behavior was coded live at the specific utterance and global level; group leader behavior was also coded globally. Interrater reliability of the coding system was assessed and preliminary validity of the coding system was examined by exploring associations between characteristics of group members and in-session group behavior. Initial reliability estimates were excellent for the specific behavior codes. Reliability of the global codes was mixed, with raters demonstrating good reliability on Support for Unhealthy Behavior, Opposition to Unhealthy Behavior, and Support for Healthy Behavior. Reliability of the group leader codes was fair to poor. Greater percent healthy talk was associated with a lower percentage of group members reporting lifetime alcohol use. The results of the current study suggest that some in-session behavior at the group level can be coded reliably via live observation and that in-session behavior at the group level is associated with alcohol use prior to attending the program. Future research is needed to explore the utility of in-session behavior in terms of predicting future behavior at the group and individual level.

Keywords: adolescent, alcohol and drug prevention, client language, group motivational enhancement

## Introduction

Significant resources have been dedicated to the prevention of alcohol and drug problems in high school students. Due to limited funding and developmental considerations, there are advantages to offering adolescent substance use prevention services in a group format (D'Amico et al., 2011). However, a recent review of brief school-based prevention programs suggests that while group-based programs do not result in iatrogenic effects, they do not yield positive effects either (Hennessy & Tanner-Smith, 2014). Given these concerns, extensive efforts have been made to identify the active ingredients in effective prevention campaigns for youth. In terms of delivery, interactive formats appear to be preferable; in terms of content, exploration of perceived norms, readiness to change, and intentions for future behavior are important (Botvin & Griffin, 2007; Cuijpers, 2002). As a client-centered, directive approach for exploring ambivalence and encouraging positive health behavior change based on the principles of motivational interviewing (Miller & Rollnick, 2002), motivational enhancement (ME) strategies often incorporate these components of effective prevention programs. For example, a community-administered ME group for at-risk adolescents demonstrated effectiveness compared to the control group on a number of outcomes including increased readiness to change over time, reduced frequency of drinking, and increased knowledge regarding alcohol and related problems (Bailey, Baker, Webster, & Lewin, 2004). School-based programs that use ME techniques appear particularly beneficial among youth (Hennessy & Tanner-Smith, 2014), but generally are offered to individuals. Further research is needed on the effectiveness of group-delivered ME in school settings.

An advantage of studying ME-based groups is the established body of literature examining in-session behavior, an important consideration in interactive group formats. Specific

statements made by individuals in support of positive health behavior change, often conceptualized as change talk (CT), is a hypothesized mechanism of action underlying motivational enhancement (Hettema, Steele, & Miller, 2005; Miller & Rollnick, 2002) that has been associated with improved outcomes in individual adult treatment samples (Apodaca & Longabaugh, 2009; Moyers et al., 2007, 2009) and nontreatment-seeking adolescents (Baer et al., 2008). However, health-promoting statements have been predominantly conceptualized and measured at the individual level; examination of such language in prevention groups is relatively new.

Although limited, previous research suggests that ME group behavior can be reliably measured (D'Amico et al., 2012; Engle, Macgowan, Wagner, & Amrhein, 2010). D'Amico and colleagues (2012) demonstrated the feasibility of taped and live coding of ME therapist integrity in groups of adolescents assigned to attend an alcohol or drug (AOD) awareness group due to a first-time AOD offense. Engle and colleagues (2010) successfully coded group commitment language, a specific form of CT, during school-based targeted group intervention for substance use problems. Greater group commitment language in later group sessions was associated with lower past 30-day marijuana use at 12-month follow up. Such research is an important prerequisite for exploring mechanisms of action in adolescent group settings.

The current study aims to build on this line of research by evaluating a strategy for measuring live in-session behavior of adolescents participating in a universal school-based drug and alcohol prevention program (Project Options; Brown, 2001). As a harm-reduction approach designed to delay the onset of substance use and prevent higher risk substance use engagement in high school, Project Options is effective at facilitating attempts to cut down/quit alcohol use in high frequency drinkers (Brown, 2001; Brown et al., 2005). Using a voluntary, self-selection

format during school lunch time, Project Options combines ME and cognitive-behavioral approaches to address perceived alcohol and marijuana norms, challenge alcohol expectancies, teach stress management, explore the advantages/disadvantages of change, explore alternative activities/reinforcers, and reinforce positive communication skills. Specific ME techniques utilized by group leaders include adoption of an empathic, non-confrontational style and facilitation of discussion via reflective statements and open-ended questions.

The current study contributes to the research literature in some important respects. First, the majority of the research on client language related to health behavior has been conducted with targeted at-risk or treatment samples (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003; Gaume et al., 2010, 2013; Moyers et al., 2007, 2009). The current study departs from the previous literature in that participants in this sample endorse a broad range of substance use involvement, from lifetime abstainers to heavy/problem users; thus, we were required to address the additional methodological challenge of coding client language for individuals that may not need to change, but rather maintain healthy behavior. Additionally, previous coding studies have utilized audio recordings and transcriptions to capture in-session behavior (e.g., Amrhein et al., 2003; Barnett et al., 2014; Engle et al., 2010; Moyer et al., 2009), yet such methods are time- and resource-intensive and may not be feasible in some settings. Of the studies reviewed, only D'Amico et al. (2012) used behavioral data from live observation; in their study, the behavior of the therapist was measured, not that of the adolescent. The current study relies exclusively on live observation for capturing in-session behavior of youth and group leaders. Finally, as mentioned above, the majority of client language research has been conducted in individual settings; the current study examines client language within a group context. As such, the goal of this study was to describe a method for capturing in-session behavior of adolescents in

prevention groups and explore the psychometric properties of that method. Drawing from the ME literature, client language during session was identified as a hypothesized mechanism of action associated with healthy decision-making around drug and alcohol use. The development and testing of such instruments is critical to the study of mechanisms of change, which can be used to improve prevention efforts and techniques.

The primary aim of the study was examine the interrater reliability of a live observation behavioral coding system capturing group-level client language in a prevention sample. A second aim of this study was to conduct a preliminary investigation of the validity of the coding system. Validity was examined by correlating self-reported alcohol use by group members with in-session group process, particularly in terms of rates of healthy and unhealthy verbal behavior. It was hypothesized that similar associations between group-level client language and substance use would be observed in a voluntary sample of high school students as those from previous studies of targeted interventions for at-risk adolescents (e.g., Baer et al., 2008; Engle et al., 2010). Specifically, greater alcohol use and weaker future intentions to cut down/quit would be associated with lower rates of healthy statements during the group session.

## **Method**

### **Participants**

The current study utilizes data from an ongoing multisite study designed to compare ME and psychoeducational control formats of Project Options and examine mediators of the prevention program effects. Participants were 425 students attending six high schools in Portland, OR, Minneapolis, MN, and Miami, FL. Sample demographics are provided in Table 1. All grades were represented and the sample was diverse in terms of race/ethnicity. A majority of

students reported lifetime alcohol use at baseline; fewer reported lifetime marijuana use. Among lifetime users, there was significant variation in terms of use and experienced problems.

### **Procedure**

Parental consent and participant verbal assent was required prior to student enrollment in the study. Students were voluntary participants and once enrolled could attend sessions on a drop-in basis; group members could change from one session to the next given the open format. All prevention sessions were conducted on-site at participating high schools during lunch periods and were capped at ten participants. Participants completed a brief baseline assessment immediately prior to their first session. One round of Project Options consisted of six sessions delivered consecutively over the course of three weeks; upon completion of the sixth session the next round began again at session one. Each session focused on different content (e.g., topic 1 dealt with perceived norms, topic 2 addressed expectancy effects, etc.) and was designed to exist as a standalone intervention. Given the drop-in nature of group, participants could attend sessions spread across multiple rounds (i.e. a student could first attend topic 4, and then return two weeks later and attend topic 2 of the next round of six sessions). At each site, two trained group leaders facilitated prevention sessions. Content of the ME and control conditions were identical; however, group leader *style* differed based on the condition. In the ME condition, group leaders were instructed to elicit more from the students by exploring ambivalence and highlighting discrepancies with the group and were trained to collaborate with the group around issues and problem-solving. In the psychoeducation control condition, group leaders were instructed to assume a more traditional teacher/expert role and engage with the group more didactically. At the end of the lunch period, students completed a rating form for the day's session prior to going to their next class.



During the prevention session, research assistants trained on the behavioral coding system rated the session for specific verbal utterances and global codes related to the session content. All behavioral coding was done live; the inability to conduct recordings, related to issues of student confidentiality for some districts, led to mutually exclusive sets of raters by site. The current study consisted of 295 group sessions. Of these sessions, 62 (21%) were coded by multiple raters. Coded sessions were evenly distributed across the six topics (i.e. of the overall sample of ratings, 18% were topic 1, 17% topic 2, 16% topic 3, 16% topic 4, 17% topic 5, and 17% topic 6; in the double-coded subset of ratings, 18% were topic 1, 16% topic 2, 19% topic 3, 13% topic 4, 19% topic 5, and 14% topic 6).

**Coding fidelity.** Raters consisted of undergraduate research assistants, graduate students, and postdoctoral fellows. In training, all study sites met via teleconference to review the coding manual to arrive at a cohesive understanding of each code. Coders practiced using audiotapes of prior sessions and discussed discrepancies among raters under supervision of the first author (BOL). Average training time for coders was ten hours. Throughout the study, individual sites held frequent coder meetings and all sites participated in monthly coding meetings to monitor drift.

## **Measures**

**Group Behavior Coding System.** Behavioral codes were largely adapted from the CLEAR Coding System (Glynn & Moyers, 2012, <http://casaa.unm.edu/download/CLEAR.pdf>), a simplified version of the Motivational Interviewing Skill Code ([MISC 2.1]; Miller, Moyers, Ernst, & Amrhein, 2008) to capture in-session client language. Coding captured in-session behavior of both student members and group leaders and was conducted at the utterance level for youth (i.e., a single unit of speech encompassing a single thought). Each participant utterance

was coded along three domains (healthy, unhealthy, and Neutral/Ambiguous). Healthy statements were those promoting reductions in and/or abstinence from alcohol and drugs or statements rejecting hazardous use and/or increases in use. Unhealthy statements promoted alcohol and drug use and/or increases in use or rejected abstinence and/or reductions in use. Neutral/Ambiguous captured statements that did not fall into the healthy or unhealthy categories. Statements were coded during the discussion section of each prevention session, when participants participated in a group dialogue about the session topic and were most likely to engage in healthy or unhealthy talk. Raters tallied the number of utterances made by group members for each of the three categories within each defined coding period. A summary variable, percent healthy talk, was calculated as  $\text{healthy}/(\text{healthy}+\text{unhealthy})$ , similar to percent CT used in MI coding studies (Miller, 2000; Moyers et al., 2009).

Raters also coded overall impressions of group behavior using a seven point Likert scale for the full session on eight criteria (Table 2). Healthy vs. unhealthy global behavior was defined similarly to the specific behavior codes above. The global group codes were selected by the study investigators at the outset of the study as hypothesized mechanisms of change related to group process. In addition to global ratings of the group, raters coded group leader behavior on five items hypothesized to be important factors of group effectiveness (Table 2). For each global code, raters were oriented to the Likert scale using anchors that provided descriptions and specific examples of session behavior that should be rated as 1, 4, or 7.

**Self-Report Measures.** Students completed a brief baseline assessment prior to the start of their first session. This assessment consisted of demographic information (i.e. age, gender, grade, race/ethnicity) and alcohol and marijuana use. Baseline lifetime alcohol use was assessed on an eight point scale from 0 = 0 to 7 = over 100 (Johnston, O'Malley, Bachman, &

Schulenberg, 2005). Due to violations of normality, this item was dichotomized into lifetime abstainers and users (0 = no, 1 = yes). Lifetime marijuana use was measured in the same fashion. Future alcohol use intentions were assessed via a single multiple choice item “Next month I will... definitely not drink (-2), probably not drink (-1), not sure (0), probably will drink (1), definitely will drink (2)” (Christiansen, Goldman, & Brown, 1995). Frequency of lifetime negative alcohol-related consequences was assessed by six items (Sadava, 1986). Each item was rated on a 10-point scale from “0” to “9 or more” resulting a total possible range of 0-54.

Frequency of lifetime marijuana-related consequences was assessed on a single item using the same 10-point scale that asked “how many times have any of the above (in reference to the six alcohol items) happened to you because of your using marijuana?” At the end of every session, students completed self-rating forms indicating whether they had consumed alcohol or not in the past month (0 = no; 1 = yes). Group level use variables were computed by averaging responses across group members. Thus, the lifetime and past 30 day use variables represent the proportion of students in the group reporting use, the negative consequences variables indicate the average number of lifetime problems experienced by group members, and higher future intentions scores indicate greater intention to drinking in the next month averaged across group members.

### **Analytic Plan**

**Coding reliability.** Two estimates were selected to assess the reliability of coding system: an intraclass coefficient (ICC) and a measure of within-session standard deviation (WSSD: Bland & Altman, 1986, 2007). The ICC provides a conservative estimate of reliability that is generalizable across raters. Due to the lack of fully-crossed design (i.e. mutually exclusive sets of raters coded different sets of sessions) of the current study, a one-way random effects ICC was selected as appropriate (Hallgren, 2012; Shrout & Fleiss, 1979). Additionally,

we report a single-measures one-way ICC, which tends to be more conservative than an average-measures ICC and allows generalization of reliability to all sessions regardless of whether sessions were double- or single-coded (Hallgren, 2012). According to Cicchetti (1994), suggested cutoff scores for ICCs are: less than .4 = poor, .4-.59 = fair, .6-.74 = good, and greater than .75 = excellent. However, as issues of low variance and restriction of range may result in ICCs that do not accurately represent interrater reliability (Hallgren, 2012), a second measure of reliability also was utilized.

The 2\*WSSD, originally reported in D'Amico et al. (2012), represents an alternative strategy of using within-session standard deviation (WSSD) to estimate the difference between raters. The 2\*WSSD squares the difference between each rating and the mean rating for a given session; thus, larger differences between raters result in proportionally larger estimates.

Although the 2\*WSSD offers a less traditional estimate of reliability, it allows for examination of the difference between raters on the units of the scale in question and addresses some of the issues of low variance/restriction of range to which the ICC is susceptible (D'Amico et al., 2012). Using a 5-point global scale, D'Amico and colleagues (2012) set benchmarks of within 1 point for global ratings and 6 points on behavioral counts as acceptable levels of agreement. Given the larger scale (7- vs. 5-points) and the added complexity of rating both group member and leader's behavior, a cutoff score of 2 points for global ratings was established for the current study; the benchmark of 6 points for behavioral counts previously set by D'Amico et al. was unchanged.

**Validity of the coding data.** Concurrent validity of the coding system was examined by correlating the group alcohol use, marijuana use, and future alcohol use intentions variables with percent healthy talk to test whether the rate of pro-healthy utterances during session was

associated with previous alcohol and marijuana experience. Similar analyses were conducted using global ratings of in-session behavior. Since the purpose of the current study was to examine the psychometric properties of a behavioral coding system effect of intervention condition was not examined directly; however, the effect of condition on in-session behavior was controlled for and semipartial correlations are reported. As groups may have reported no alcohol-related problems due to lack of lifetime alcohol use across group members (as opposed to groups reporting lifetime drinking with no negative consequences), associations among in-session behavior and alcohol-related problems were assessed within groups reporting some group lifetime alcohol use only (i.e. groups where all members were lifetime abstainers were excluded). The same strategy was used for marijuana-related problems.

## **Results**

### **Behavioral Coding Reliability**

Initial reliability estimates of the 16 behavioral codes from the 62 sessions coded by multiple raters are provided in Table 3. The reliability of the specific behavior codes (healthy, unhealthy, neutral) was excellent. Reliability of the percent healthy talk variable calculated from the healthy and unhealthy behavior codes was good (ICC = 0.695, 2\*WSSD = 0.16). In terms of the global student ratings, good reliability was found on the Support for Unhealthy/Risky Behavior, Opposition to Unhealthy/Risky Behavior, and Support for Healthy/Low Risk Behavior codes. The remaining student global codes yielded poor ICC interrater reliability indicators. The 2\*WSSD score was within the acceptable range of agreement for all student global codes. For Opposition to Healthy/Low Risk Behavior and both Change Toward codes, restricted variance may have contributed to an unduly conservative estimate of the ICC. Using the ICC, reliability for the global leader codes was fair for Open-ended Questions, Reflective Statements, and Style

of Interaction and poor for Warmth/Nonjudgment and Script Adherence. The 2\*WSSD score was within the acceptable range of agreement for all global leader codes. Script Adherence demonstrated restricted variance.

### **Description of In-session Behavior**

Based on the overall sample of 295 sessions, group members provided statements supporting healthy behavior more often than unhealthy behavior in sessions as evidenced by a mean (SD) of 65.5% (29.2) percent healthy talk (Table 3); however, there was a large range of behavior, with some groups providing no healthy talk relative to unhealthy talk and other groups providing all healthy talk and no unhealthy talk. At the global level, groups scored at the lower end of the scales for codes capturing substance-specific discussion, accurately reflecting that not all session content was devoted to explicit discussion of substance use (e.g., dealing with stress, positive communication). When scores of each global code were considered relative to others, groups tended to endorse healthier behavior relative to unhealthy behavior. For example, groups obtained higher ratings on Support for Healthy/Low Risk Behavior relative to Opposition to Healthy/Low Risk Behavior,  $t(285) = 15.41, p < .01$ , or Support for Unhealthy/Risky Behavior,  $t(285) = 9.34, p < .01$ . Across sessions, group members were rated on the higher end of the Engagement/Interest and Responsiveness to Leaders scales (Table 3).

In terms of group leader behavior, leaders were rated as more frequently using open-ended than closed questions and reflective statements than other types of responses (e.g., providing information or advice) to member statements. Group leaders were rated as slightly more collaborative than instructional and received high scores on Warmth/Nonjudgment. Group leaders consistently were rated as having high levels of script adherence.

### **Validation of Behavioral Codes**

As adequate reliability is a prerequisite for validity (Popham, 1981), only substance-related behavioral codes with acceptable levels of reliability on both reliability estimates (i.e. ICC greater than .4 and 2\*WSSD less than 2) were validated against substance use behavior. These variables included: percent healthy talk, Support for Unhealthy Behavior, Opposition to Unhealthy Behavior, and Support for Healthy Behavior. For all analyses, the effect of intervention condition was statistically controlled; results are presented with this effect covaried out. Percent healthy talk during session and the proportion of group members reporting lifetime alcohol use were negatively correlated,  $r = -0.23$ ,  $p < .01$ , such that a lower rate of healthy talk was associated with a greater percentage of lifetime users. Global Support for Unhealthy Behavior was significantly and positively associated with lifetime alcohol use,  $r = 0.22$ ,  $p < .001$ , and past 30 day alcohol use,  $r = 0.12$ ,  $p = .04$ . None of the coding variables were significantly associated with group alcohol use intentions for the next month. Within groups reporting at least some lifetime alcohol use, Support of Unhealthy Behavior, Opposition to Unhealthy Behavior, and Support for Healthy Behavior were associated with greater lifetime alcohol-related problems,  $r = 0.15$ ,  $p = .03$ ;  $r = 0.14$ ,  $p = .03$ ; and  $r = .16$ ,  $p = .02$ , respectively. None of the in-session behaviors were associated with marijuana-related problems.

### **Discussion**

The current study contributes to the limited research describing the measurement of adolescent verbal behavior during group sessions. We found that multiple categories of in-session behavior at the group level can be coded reliably with relatively minimal training (~10 hours per coder). Using the 2\*WSSD estimate and a cutoff of two points on a seven-point scale, raters were able to reliably capture group member behavior at the specific utterance and global level, as well as group leader global behavior. Using a more traditional estimate of interrater

reliability, the ICC, raters demonstrated good reliability for specific group behavior (i.e. percent healthy talk) and three substance-related global codes (Support for Unhealthy Behavior, Opposition to Unhealthy Behavior, and Support for Healthy Behavior). According to the ICC, raters demonstrated poor reliability on group engagement/interest and responsiveness codes [and](#) fair reliability on a number of group leader codes (Use of Open-ended Statements, Use of Reflective Statements, and Style of Interaction). It should be acknowledged that some differences in the level of reliability were observed based on whether one uses the 2\*WSSD or ICC estimates. For the current study, neither estimate was given priority; rather both were utilized in order to mitigate the shortcomings of each estimate alone. However, for future studies, one may opt to consider one estimate as primary based on one's goals. For example, use of the ICC allows for comparison to other coding systems which may be important in terms of replicability. Use of the 2\*WSSD allows for quantification of the difference between raters in scale units (D'Amico et al., 2012) which may be more useful for training or clinical purposes.

Beyond demonstrating the reliability of many of the codes in the current coding scheme, the current study also found that high school students voluntarily attending a prevention program provide more healthy than unhealthy verbal statements in group. As research suggests that adolescents are more likely to internalize anti-alcohol messages from peers than pro-alcohol (Teunissen et al., 2012), further research is needed to better understand the effects of peer influence in group prevention versus treatment settings. Regarding initial tests of validity, in-session comments supportive of healthy behavior by participants were negatively associated with baseline substance use history at the group level, partially supporting hypotheses. However, the hypothesized relationships between behavioral intentions in the future and in-session behavior



were not observed. This may be due to the fact that intentions were measured at the group level rather than the individual level.

Alcohol and drug treatment services are frequently offered in a group format, particularly for adolescents (see D'Amico et al., 2011). However, mechanisms of ME-based interventions have largely been studied in individual settings. This study expands the limited literature showing that client language can be reliably assessed at the group-level, allowing for further examination of the effectiveness of ME as a group process. Based on theory (e.g., Hettema et al., 2005) and empirical research (e.g., Amrhein et al., 2003, Barnett et al., 2014; Moyers et al., 2007, 2009), client language related to behavior change appears central when exploring the mechanisms of change underlying ME treatment strategies. Preliminary evidence suggests that some measures of client language are important in predicting substance use change in at-risk youth seeking services (Baer et al., 2008). We attempted to capture a measure of client language related to healthy behavior appropriate to a universal prevention program as opposed to previous work in treatment settings (all adult studies) or targeted intervention with at-risk adolescents (e.g., D'Amico et al., 2012; Engle et al., 2010). Specifically, the target behavior was not non-use vs. use, but rather healthy behavior (including continued non-use, lower risk use, and reductions in use) vs. unhealthy behavior (maintaining higher risk use and increases in use). This harm-reduction healthy talk definition requires further validation and comparison with more traditional CT definitions used in treatment samples.

The current study applied a behavioral coding system within a school-based implementation study. While rigorous coding methods utilizing audio recordings and transcriptions yield important and useful in-depth data (e.g., Amrhein et al., 2003; Barnett et al., 2014; Moyer et al., 2009), they often are not practical or even feasible in many settings. Not

only does the current study indicate that in-session data can be reliably captured through live observation, but it also offers a technique that clinicians can use in real-time to evaluate group behavior. For example, given the associations observed among baseline use variables (which providers can have access to prior to a group session) and in-session behavior, providers may prioritize the reinforcement of certain types of behavior over others based on composition of the group. However, the effects of live observation versus recordings also may provide interesting data in the future. There are additional limitations of the current study that bear consideration. The sample of raters was comprised of mutually exclusive groups of raters which limited the reliability estimates that could be utilized (e.g., rendered a two-way ICC inappropriate) and may have influenced the reliability of the system. Also, some of the global behavior codes demonstrated poor reliability according to the ICC estimate, particularly among the group leader codes. One possible explanation for the weaker reliability on some of the global codes could be that the demands of the live coding were greater than in previous studies (e.g., the MISC 2.1 recommends coders make 2-3 passes for each session, thereby rating client and provider behaviors and/or global and specific behaviors separately). As reliability is sample-specific (Ebel, 1979), perhaps reliability of this coding system could be improved upon in the future by having separate coders rate group member behavior and group leader behavior. Due to the focus on behavioral coding of groups, we only examined the associations among in-session behavior and group-level use variables; different relations may be observed at the individual level.

This study represents an expansion of previous research on ME and client language to the prevention area, something particularly important in adolescent samples (D'Amico et al., 2011). Ultimately, the current study provides initial evidence that in-session behavior of ME-based groups of nonclinical adolescents can be reliably measured in real-world settings without

requiring prohibitive levels of training and resources. It also highlights a number of exciting areas of research that may shed light on important processes of adolescent groups with future empirical study, particularly in terms of group leader behavior and interactions between group members and group leaders. Research suggests that providers play an important role in the expression of types of client language (e.g., Glynn & Moyers, 2010), and future studies should explore what provider behaviors are associated with changes in group behavior for youth. .

Additionally, this study offers preliminary evidence that the substance use experience of group members influences group language; further research is needed to better understand such relationships and how to best train providers to adapt to such group dynamics. As ME groups

are relatively understudied, research on mechanisms of action will require an integration of the evidence from individual research with research and theory on group process and peer influence. From a prevention standpoint, the utility of in-session behavior to predict future behavior needs to be explored, including not only healthy behavior *change*, but also *maintenance* of pre-existing healthy behavior.

### **Acknowledgments**

The authors declare that they have no conflict of interest. This study was funded by the National Institute on Alcohol Abuse and Alcoholism (grant number 5R01AA012171-13). The authors would like to acknowledge Drs. Sandra Brown (PI of the parent study), Eric Wagner (Co-I), and Ken Winters (Co-I) and thank the many research assistants who made this study possible.

### References

- 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*, 862–878.
- Apodaca, T.R., & Longabaugh, R. (2009). Mechanisms of action in motivational interviewing: A review of the evidence. *Addiction, 104*, 705–715.
- Baer, J.S., Beadnell, B., Garrett, S.B., Hartzler, B., Wells, E., & Peterson, P.L. (2008). Adolescent change language within a brief motivational intervention and substance use outcomes. *Psychology of Addictive Behaviors, 22*, 570–575.
- Bailey, K.A., Baker, A.L., Webster, R.A., & Lewin, T.J. (2004). Pilot randomized controlled trial of a brief alcohol intervention group for adolescents. *Drug and Alcohol Review, 23*, 157-166.
- Barnett, E., Moyers, T.B., Sussman, S., Smith, C., Rohrbach, L.A., Sun, P., & Sprujit-Metz, D. (2014). From counselor skill to decreased marijuana use: Does change talk matter? *Journal of Substance Abuse Treatment, 46*, 498-505.
- Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307– 310.
- Bland, J.M., & Altman, D.G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics, 17*, 571–582.
- Botvin, G.J. & Griffin, K.W. (2007). School-based programmes to prevent alcohol, tobacco and other drug use. *International Review of Psychiatry, 19*, 607-615.

- Brown, S.A. (2001). Facilitating change for adolescent alcohol problems: A multiple options approach. In E.F. Wagner, & H.B. Waldron (Eds.), *Innovations in adolescent substance abuse intervention* (pp. 169–187). Oxford: Elsevier Science.
- Brown, S.A., Anderson, K.G., Schulte, M.T., Sintov, N.D., & Frissell, K.C. (2005). Facilitating youth self-change through school-based intervention. *Addictive Behaviors, 30*, 1797-1810.
- Christiansen, B.A., Goldman, M.S., & Brown, S.A. (1995). Alcohol Expectancy Questionnaire: Adolescent form, in *NIAAA Assessing Alcohol Problems: A Guide for Clinicians and Researchers*, p. 223-228. National Institute of Alcohol Abuse and Alcoholism: Rockville, MD.
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.
- Cuijpers, P. (2002). Effective ingredients of school-based drug prevention programs: A systematic review. *Addictive Behaviors, 27*, 1009-1023.
- D’Amico, E.J., Feldstein Ewing, S.W., Engle, B., Hunter, S.B., Osilla, K.C., & Bryan, A. (2011). Group alcohol and drug treatment. In (Eds. Naar-King, S. & Suarez, M.) *Motivational Interviewing with adolescents and young adults*, pp. 151-157. New York, NY: Guilford Press.
- D’Amico, E.J., Osilla, K.C., Miles, J.N.V., Ewing, B., Sullivan, K., Katz, K., & Hunter, S.B. (2012). Assessing motivational interviewing integrity for group interventions with adolescents. *Psychology of Addictive Behaviors, 26*, 994-1000.

Ebel, R.L. (1979). *Essentials of educational measurement (3rd ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Engle, B., Macgowan, M.J., Wagner, E.F., & Amrhein, P.C. (2010). Markers of marijuana use outcomes with in adolescent substance abuse group treatment. *Research on Social Work Practice, 20*, 271-282.

Gaume, J., Bertholet, N., Faouzi, M., Gmel, G., & Daeppen, J.B. (2010). Counselor motivational interviewing skills and young adult change talk articulation during brief motivational interventions. *Journal of Substance Abuse Treatment, 39*, 272–281.

Gaume, J., Bertholet, N., Faouzi, M., Gmel, G., & Daeppen, J.B. (2013). Does change talk during brief motivational interventions with young men predict change in alcohol use? *Journal of Substance Abuse Treatment, 44*, 177–185.

Glynn, L.H. & Moyers, T.B. (2010). Chasing change talk: The clinician’s role in evoking client language about changing. *Journal of Substance Abuse Treatment, 39*, 65-70.

Glynn, L.H. & Moyers, T.B. (2012). *Manual for the Client Language Easy Rating (CLEAR) Coding System*. Retrieved from <http://casaa.unm.edu/download/CLEAR.pdf>.

Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23-34.

Hennessy, E.A. & Tanner-Smith, E.E. (2014). Effectiveness of brief school-based interventions for adolescents: A meta-analysis of alcohol use prevention programs. *Prevention Science, online publication 10/8/14*.

Hettema, J., Steele, J., Miller, W.R. (2005). Motivational interviewing. *Annual Review of Clinical Psychology, 1*, 91–111.

- Johnston, L.D., O'Malley, P.M., Bachman, J.G., & Schulenberg, J.E. (2006). Monitoring the Future national survey results on drug use, 1975-2005. *Volume I: Secondary school students (NIH Publication No. 06-5883)*. National Institute on Drug Abuse: Bethesda, MD.
- Miller, W. R. (2000). *The Motivational Interviewing Skill Code (MISC) manual*. Retrieved from <http://casaa.unm.edu/download/misc1.pdf>
- Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). *Manual for the Motivational Interviewing Skill Code Version 2.1 (MISC)*. Retrieved from <http://casaa.unm.edu/download/misc.pdf>.
- Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*, 2nd ed. New York: Guilford.
- Moyers, T.B., Martin, T., Christopher, P.J., Houck, J.M., Tonigan, J.S., & Amrhein, P.C. (2007). Client Language as a mediator of motivational interviewing efficacy: Where is the evidence? *Alcoholism: Clinical and Experimental Research*, 31, 40S–47S.
- Moyers, T.B., Martin, T., Houck, J.M., Christopher, P.J., Tonigan, J.S. (2009). From in-session behaviors to drinking outcomes: A causal chain for motivational interviewing. *Journal of Consulting and Clinical Psychology*, 77, 1113-1124.
- [Popham, W.J. \(1981\). \*Modern educational measurement\*. Englewood Cliffs, NJ: Prentice-Hall.](#)
- Sadava, S.W. (1986). Alcohol consumption and alcohol problems: Gender differences. *Bulletin of the Society of Psychologists in Addictive Behaviors*, 5, 67-73.
- Shrout, P.E., Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.



- Teunissen, H.A., Spijkerman, R., Prinstein, M.J., Cohen, G.L., Engels, R.C.M.E., & Scholte, R.H.J. (2012). Adolescents' conformity to their peers' pro-alcohol and anti-alcohol norms: The power of popularity. *Alcoholism: Clinical and Experimental Research*, 36, 1257-1267.
- Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

Table 1. Sample Demographics

<b>Group Variable</b>	<b>M</b>	<b>SD</b>	<b>%</b>	<b>N</b>
Age (years)	16.0	1.4		
Gender (female)			61.7	261
Grade				
9th			28.4	120
10th			18.7	79
11th			23.0	97
12th			29.9	126
Race				
White			42.1	179
Asian			5.2	22
Black/African American			24.9	106
Mixed Race			8.9	38
Other			18.8	80
Ethnicity (% Latino/Hispanic)			32.2	137
Lifetime Alcohol User (% yes)			52.7	224
Past 30 Day Drinking Days	1.3	2.7		
Lifetime Alcohol Problems*	5.1	9.0		
Alcohol Use Intentions	-1.3	1.0		
Lifetime Marijuana User (% yes)			40.5	172
Lifetime Marijuana Problems*	1.9	2.7		

\* represents frequency of problems experienced among lifetime users only. An intentions score of -1 corresponds to a response of “I will probably not drink” in the next month, alcohol problems could range from 0-54, marijuana problems from 0-9.

Table 2. Description of Global Codes and Scale Anchors

	<b>Global Student Codes</b>	<b>Likert scale anchors</b>	
		<b>1</b>	<b>7</b>
1	How often did the students make statements supportive of unhealthy/risky behavior	Not at all	Frequently
2	How often did the students make statements opposed to unhealthy/risky behavior	Not at all	Frequently
3	How often did the students make statements supportive of healthy/low-risk behavior	Not at all	Frequently
4	How often did the students make statements opposed to healthy/low-risk behavior	Not at all	Frequently
5	How often did the students make statements for changing toward unhealthy/risky behavior	Not at all	Frequently
6	How often did the students make statements for changing toward healthy/low-risk behavior	Not at all	Frequently
7	How engaged/involved/interested were students in the session	Not at all	Very Much
8	How responsive were students to the leaders' questions/comments	Not at all	Very Much
	<b>Global Leader Codes</b>		
1	How often did the leader(s) use open-ended questions	Not at all	Frequently
2	How often did the leader(s) make reflective statements	Not at all	Frequently
3	Please rate the style of interaction between leaders(s) and participants	Instructional	Collaborative
4	To what degree were the leader(s) warm and nonjudgmental	Very Cool	Very Warm
5	How closely did the leader(s) follow the script	Not at all	Completely

Table 3. Reliability and mean (SD) ratings on behavioral coding system

Code		2*WSS			
		D	ICC	Mean	SD
<b>Behavior</b>	Unhealthy statements	0.80	0.798	1.73	2.45
	Healthy statements	1.84	0.758	3.41	4.29
<b>r Codes</b>	Neutral/ambiguous statements	3.36	0.897	6.36	9.44
	Support for unhealthy/risky behavior	0.60	0.65	1.80	1.17
<b>Student</b>	Opposition to unhealthy/risky behavior	1.48	0.572	2.69	1.77
	Support for healthy/low risk behavior	1.30	0.684	3.01	1.91
	Opposition to healthy/low risk behavior	0.53	0.303	1.30	0.70
	Change toward unhealthy/risky behavior	0.16	-0.024	1.04	0.33
<b>Global Ratings</b>	Change toward healthy/low risk behavior		0.165		
	Engagement/Interest	1.32	0.369	5.36	1.07
	Responsiveness to leaders	1.46	0.288	5.34	1.21
	Use of open-ended questions	1.81	0.399	5.04	1.66
<b>Leaders</b>	Use of reflective statements	1.65	0.445	4.62	1.84
	Style of interaction	1.44	0.437	4.73	1.51
<b>Global Ratings</b>	Warmth/Nonjudgment	1.46	-0.090	6.16	0.94
	Script adherence	0.74	0.063	6.13	0.72

*Note:* Reliability estimates based on the 62 sessions coded by multiple raters, the mean (SD)

scores based on the full 295 sessions.