

Children's Mercy Kansas City

SHARE @ Children's Mercy

Manuscripts, Articles, Book Chapters and Other Papers

11-1-2016

The Alcohol Intervention Mechanisms Scale (AIMS): Preliminary Reliability and Validity of a Common Factor Observational Rating Measure.

M Magill

Timothy R. Apodaca
Children's Mercy Hospital

Justin Walthers

Jacques Gaume

Ayla Durst

See next page for additional authors

Let us know how access to this publication benefits you

Follow this and additional works at: <https://scholarlyexchange.childrensmercy.org/papers>



Part of the [Behavioral Medicine Commons](#), [Cognitive Behavioral Therapy Commons](#), and the [Substance Abuse and Addiction Commons](#)

Recommended Citation

Magill M, Apodaca TR, Walthers J, et al. The Alcohol Intervention Mechanisms Scale (AIMS): Preliminary Reliability and Validity of a Common Factor Observational Rating Measure. *J Subst Abuse Treat.* 2016;70:28-34. doi:10.1016/j.jsat.2016.07.012

This Article is brought to you for free and open access by SHARE @ Children's Mercy. It has been accepted for inclusion in Manuscripts, Articles, Book Chapters and Other Papers by an authorized administrator of SHARE @ Children's Mercy. For more information, please contact hlsteel@cmh.edu.

Creator(s)

M Magill, Timothy R. Apodaca, Justin Walthers, Jacques Gaume, Ayla Durst, Richard Longabaugh, Robert L. Stout, and Kathleen M. Carroll



Published in final edited form as:

J Subst Abuse Treat. 2016 November ; 70: 28–34. doi:10.1016/j.jsat.2016.07.012.

The Alcohol Intervention Mechanisms Scale (AIMS): Preliminary Reliability and Validity of a Common Factor Observational Rating Measure

M. Magill¹, T.R. Apodaca^{2,3}, J. Walthers¹, J. Gaume^{1,4}, A. Durst¹, R. Longabaugh¹, R.L. Stout^{1,5}, and K.M. Carroll⁶

¹Center for Alcohol and Addiction Studies, Brown University, Providence, RI, USA ²Children's Mercy Kansas City, USA ³University of Missouri-Kansas City School of Medicine, USA ⁴Lausanne University Hospital, Lausanne, Switzerland ⁵Pacific Institute for Research and Evaluation, Providence, RI, USA ⁶Yale University School of Medicine, New Haven, CT, USA

Abstract

The present work provides an overview, and pilot reliability and validity for the Alcohol Intervention Mechanisms Scale (AIMS). The AIMS measures therapist interventions that occur broadly across modalities of behavioral treatment for alcohol use disorder. It was developed based on identified commonalities in the *function* rather than *content* of therapist interventions in observed therapy sessions, as well as from existing observer rating systems. In the AIMS, the primary function areas are: Explore (four behavior count codes), Teach (five behavior count codes), and Connect (three behavior count codes). Therapist behavior counts provide a frequency rating of occurrence (i.e., adherence). The three functions (Explore, Teach, Connect) are then rated on global skillfulness, which provides a quality valence (i.e., competence) to the entire session. In the present study, three independent raters received roughly 30 hours of training on the use of the AIMS by the first author. Data were a sample of therapy session audio files from a Project MATCH clinical research site. Reliability results showed generally good performance for the measure. Specifically, 2-way mixed Intraclass Coefficients were 'excellent', ranging from .94 to .99 for function summary scores, while Prevalence-Adjusted, Bias-Adjusted Kappa for global skillfulness measures were in the 'fair' to 'moderate' range ($k = .36$ to $.40$). Internal consistency reliability was acceptable, as were preliminary factor models by behavioral treatment function (i.e., Explore, Teach, Connect). However, confirmatory fit for the subsequent three factor model was poor. In concurrent validity analyses, AIMS summary and skillfulness scores showed associations with relevant Project MATCH criterion measures (i.e., MATCH Tape Rating Scale) that were consistent with expectations. The AIMS is a promising and reliable observational measure of three proposed common functions of behavioral alcohol treatment.

Correspondence: Molly Magill, PhD, BoxG-S121-5, Providence, RI 02912, 401-863-6557, molly_magill@brown.edu.

The contents of this manuscript are the responsibility of the authors and do not represent official positions of the National Institutes of Health or the United States Government.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Alcohol; Common Factors; Process Research; Project MATCH; Psychometrics

Introduction

Why do nominally and theoretically distinct treatments for alcohol and other drug use disorders typically perform similarly well in efficacy trials? Perhaps the most striking examples are findings from two large-scale studies - Project MATCH (1997; 1998) and the United Kingdom Alcohol Treatment Trial (UKATT Research Team, 2008). In each case, results failed to show significant differential efficacy between very different methods of treatment. In MATCH, each behavioral modality (Cognitive-Behavioral Therapy [CBT], Motivational Enhancement Therapy [MET] and Twelve-Step Facilitation [TSF]) demonstrated similar improvements in the percent of days abstinent and in the number of drinks per drinking day up to 15 months post-treatment (Project MATCH Research Group, 1997). Further, of the more than 20 causal chains hypothesized to mediate the proposed modality-specific matching effects, most failed to reach statistical significance (Longabaugh & Wirtz, 2001). In the UKATT study, MET was compared to an integration of CBT and community reinforcement and results were equivalent 12 months later. Consistent with MATCH, the large majority of matching hypotheses were not supported (UKATT Research Team, 2008). The phenomenon, which also applies broadly to the field of psychotherapy (see e.g., Wampold et al., 1997), has been called the *Dodo Bird Effect*. The metaphor is based on a character in Alice's Adventures in Wonderland who claimed: "*Everybody has won, and all must have prizes!*" The ubiquity of the Dodo Bird Effect has led some to argue the importance of examining treatment process (i.e., active ingredients and mechanisms of change) as a way to better understand how behavioral treatments are working (Kazdin, 2007; Longabaugh, 2007; Longabaugh & Magill, 2011; Morgenstern & McKay, 2007).

How do behavioral treatments work? Studies often fail to validate modality-specific factors

Tests of statistical mediation with modality-specific variables is a common approach to examining how treatments work, and the addictions field has experienced substantial growth in this type of research. In early work, Morgenstern and Longabaugh (2000) examined coping skills as a mediator of CBT outcomes across 10 efficacy trials and found little evidence for the hypothesized mechanism. Recent research has been more promising (Kiluk, Nich, Babuscio, & Carroll, 2010; Petry, Litt, Kadden, & Ledgerwood, 2007), but the evidence remains mixed (see e.g., Litt, Kadden, Cooney, & Kabela, 2003; Litt, Kadden, & Stephens, 2005). Tests of causal process in MET/Motivational Interviewing (MI) have shown more convergent findings, but this may be due to an emphasis on mediation analyses within condition as opposed to examining mechanisms in contrast to another behavioral treatment. A recent meta-analysis on 12 MI process studies demonstrated partial support for the hypothesis that MI operates through its proposed key mechanisms, change and sustain talk, and these mechanisms are influenced by therapist behaviors consistent with MI principles (Magill et al., 2014). Another recent review found support for some MI-specific processes (e.g., change talk, discrepancy) and not others (e.g., motivation; Apodaca & Longabaugh, 2009). So while there is some support for the MI process model, between-

treatment comparisons are required to rule out the possibility that these variables are operative in other treatments as well. Finally, research has supported that TSF exerts at least a portion of its effects on drinking through common rather than treatment-specific processes. In a review of 19 studies of Alcoholics Anonymous (AA) and 12-step related treatments, Kelly, Magill and Stout (2009) found the most compelling evidence for coping, motivation, and self-efficacy, as mediators of alcohol use reduction. Moreover, Forchimes and Tonigan (2008) conducted a meta-analysis of 11 studies and found modest support for self-efficacy as a mediator of AA effects. In sum, in these three, frequently-utilized behavioral treatments, the support for modality-specific factors appears more limited than that for common factors.

How do we explain the Dodo Bird Effect? Three perspectives

Research to date has not converged on a modality-specific or common factor causal process model for behavioral addictions treatment. A shared conceptual framework could guide future research efforts. We propose three potential ways to explain non-differential efficacy in controlled outcome trials of evidence-based treatments. First, it is possible that the treatments have unique ingredients and mechanisms, but these processes have equal efficacy. In other words, *there are multiple viable routes to the same outcome*. Second, the treatments might have unique key ingredients, but these ingredients do not surpass the effects of shared client mechanisms of change (e.g., motivation, self-efficacy, network support). Thus, the field can acknowledge *there are both modality-specific and common factors*. Third, the most powerful ingredients and mechanisms could be those the treatments share rather than those that make them unique. That is, *behavioral interventions work through common factor variables*. Each perspective has advocates in the literature (Hofman & Barlow, 2014; Laska, Gurman, & Wampold, 2014; Prochaska & DiClemente, 1986), but meta-analytic studies have provided compelling support for a common factor framework for psychotherapy (Wampold, 2001) and for alcohol treatment (Imel, Wampold, Miller, & Fleming, 2008). In the present work, we emphasize this third perspective, an important and understudied topic in the addictions.

Purpose

The present work provides an overview, and pilot reliability and validity for a novel observational rating system of common therapeutic factors in behavioral alcohol use disorder treatment, the Alcohol Intervention Mechanisms Scale (AIMS). The AIMS was designed to measure therapist interventions that occur broadly across treatment modalities, and it was developed based on identified commonalities in the *function* rather than *content* of therapist interventions in observed therapy sessions, as well as from existing observer rating systems (e.g., Miller, Moyers, Ernst, & Amrhein 2003; 2008; Nuro et al., 2001). The three AIMS therapeutic functions are to: Explore, Teach, and Connect. The goal of the measure is to succinctly capture the exploratory and didactic nature of behavioral alcohol treatment while also measuring the relational/interpersonal capacities of the therapist and/or therapy. In the current study, we used a sample of therapy session audio files from a Project MATCH clinical research site, which enabled an examination of measure psychometrics and purported common factor processes across three evidence-based treatments (CBT, MET, TSF). We pursued the follow research aims:

1. Report inter-rater and internal consistency reliability for the measure;
2. Conduct confirmatory factor analysis of the proposed therapy functions;
3. Test correlations with convergent criteria from existing Project MATCH data.

Method

The AIMS

Overview

Conceptual model for the AIMS: As noted above, the AIMS was developed based on the notion that behavioral interventions may look different, but function the same. For example, one therapist might inquire about “triggers” for substance use while another might ask about “denial” of use severity, but both therapist are exploring barriers to initiation of abstinence or risks for relapse. The function is to explore change. Also, when therapists provide didactic instruction, the *content* may relate to a variety of topics such as cognitive copings skills, normative alcohol use patterns, or 12-step philosophy, but the *function* is to provide information, to teach, or to advise. In both examples, the intermediate outcome is knowledge within the client while the exact content of that knowledge may differ.

Structure of the AIMS: In the AIMS, the primary functions are: Explore (four codes), Teach (five codes), and Connect (three codes); see Table One. These behavior count codes are intended to be ‘quality-neutral’, providing information about frequency of occurrence (i.e., adherence). Each of the three functions is then rated on a five-point, ordinal skillfulness scale, which provides a quality valence to the entire session (i.e., competence). Here, raters are instructed to begin at a score of three and lower or raise their scores based on quality descriptors. Finally, there are three codes: confront/challenge, general information, and neutral/facilitate that do not fall under a primary function category.

Development of the AIMS: The AIMS was developed inductively via coded behavioral alcohol treatment dialogues, but was also informed deductively using indicators in existing observational process measures. For example, the Yale Adherence and Competence Scale (Carroll et al., 2000; Nuro et al., 2005), UKATT Process Rating Manual (Tober, Clyne, Finnegan, Farrin, & Russel, 2008), and the Motivational Interviewing Skill Code (Miller et al., 2003; 2008) are all examples of rating systems that target specific modalities, and that could be mined for items differing in content but overlapping in function.

Rater training—For the present psychometric report, three bachelor’s level raters received roughly 30 hours of training from the first author. Rater training followed standard procedures, including the use of audio-recorded pilot sessions from a training library ($N=7$). These sessions have exemplar ratings of therapist codes with narrative justification. Observational rater training involved three phases: 1) didactic overview, including treatment- and coding-related readings (i.e., Kadden et al., 1992; Magill & Apodaca, 2011; Miller, Zweban, DiClemente, & Rychtarik, 1992; Nowinski, Baker, & Carroll, 1992), 2) group coding practice with corrective feedback, and 3) individual coding practice with group

corrective feedback. Rater proficiency was defined by Intraclass Correlation Coefficient (ICC) agreement with training library exemplar ratings (i.e., ICC = .75 or above; Cicchetti, 1994). Weekly group sessions were held throughout the course of the study to prevent rater drift. Finally, observational raters were masked to study aims and participant outcomes.

Study sample and session selection

Observational rating data were derived from a sample of session files from a Northeast, Project MATCH aftercare site. Project MATCH (1997) tested 21 matching variables, across three multi-session, alcohol treatments (CBT, TSF, MET) at 10 research sites among 1,726 participants with alcohol use disorders. The study demonstrated significant main effects, across treatment conditions, over follow-up (Project MATCH Research Group, 1997; 1998). Participants were treatment-seeking adults meeting DSM III-R criteria for alcohol abuse or dependence. Of the original site sample ($N = 168$), session data were available for 89.9% of participants ($N = 151$). Of these cases, recorded treatment sessions were available for 99.3% ($N = 150$). Observational data were collected on four treatment sessions per condition (i.e., first through third and final), consistent with methods by Karno, Longabaugh, and Herbeck (2010). Further, we selected only those cases where at least three sessions were available (final $N = 126$; 106 four-session and 20 three-session cases; CBT = 46; TSF = 42; MET = 38). Sessions in this sample were 90 minutes in length on average ($SD = 13.00$), and there were no systematic differences in session length by condition. Participants in this sample were 45 years old on average ($SD = 13.3$), majority male (69.8%) and Caucasian (94%). The majority of participants were employed (64.2%), unmarried (59.5%), and their average years of education was 13 ($SD = 2.1$). This was a primarily alcohol dependent sample (69.6%).

Study treatments

The selected sample and sessions enabled an examination of common factor processes across the three behavioral treatments tested in MATCH. CBT and TSF involved 12 weekly sessions, while MET included four sessions, conducted at the first, second, sixth, and twelfth weeks of treatment. Each treatment had a well-specified theoretical model and corresponding manualized protocol, which we describe briefly here. First, CBT was based on a social learning model with intervention strategies targeting prescribed coping activities related to internal and external risks for relapse (e.g., managing urges/cravings, managing negative affective states, drink refusal skills, social skills training; see Kadden et al., 1992). Second, TSF was based on a disease framework and focused on involvement in Alcoholics Anonymous prescribed coping activities (e.g., acceptance of disease, meeting attendance, sponsorship, engaging in the 12-steps; see Nowinski, Baker, & Carroll, 1992). Third, MET was grounded in a theoretical integration of motivational psychology and client-centered therapy and emphasized therapeutic skills that activate client internal capacities for change (e.g., efficacy support, exploration of ambivalence, personalized feedback on alcohol use, change planning; see Miller et al., 1992). Project MATCH achieved high treatment adherence, integrity, as well as discriminant validity (Carroll et al., 1998).

Measurement

Convergent validity measures—The present study used one criterion indicator for each of the three AIMS functions (i.e., Explore, Teach, Connect), all of which were collected as an aspect of the original Project MATCH process and fidelity assessment. Specifically, criterion measures were derived from the MATCH Tape Rating Scale (MTRS), which served as the foundation for the now commonly used Yale Adherence and Competence Scale (Carroll et al., 2000; Nuro et al., 2005). The MTRS was developed to assess treatment fidelity and discriminability, and includes three treatment-specific subscales (i.e., CBT, TSF, MET), and two non-specific subscales (i.e., structure, general support). Raters mark counts of observed behaviors, and these counts are recoded to a five-point “extensiveness” scale. MTRS convergent criteria were selected from among ‘non-specific’ items. For therapist Explore, the MTRS item Depth of Exploration, defined as: “...the degree to which the therapist encouraged depth of exploration rather than shallowness”, was used. For therapist Teach, MTRS Advice Giving, defined as: “...the degree to which the therapist provides specific, concrete advice to the patient”, was used. The therapist Connect criterion was the MTRS item, Empathy or “...the degree to which the therapist responds empathetically to the patient”. Project MATCH collected MTRS ratings at sessions two and six. The current study thus reports primarily session two data to allow comparison between Project MATCH within-treatment process data and Project AIM observational rating data.

Data-Analysis

Analyses for the current psychometric report targeted examination of the inter-rater and internal consistency reliability as well as the factorial and convergent validity of the AIMS. All analyses, with the exception of Confirmatory Factor Analysis (CFA), were performed in SPSS Version 22.0 (IBM Corporation). For project inter-rater reliability, a random sample of session files ($N = 47$) was double-coded and analyses were conducted in three-month increments over the course of the study. Analyses were specified as two-way mixed effects (rater as random; measure as fixed), single measure, Intraclass Correlation Coefficients [ICC]; McGraw & Wong, 1996). For ordinal skillfulness measures, Prevalence-Adjusted, Bias-Adjusted Kappa (PABAK) values were examined. Here, ratings tended to cluster towards the middle score, which have been shown to result in misleadingly low values for Cohen’s kappa (Byrt, Bishop, & Carlin, 1993; Hallgren, 2012). Internal consistency analyses were completed with Cronbach’s alpha. Finally, Pearson bivariate correlations assessed convergent validity between AIMS summary scores and skillfulness measures and Project MATCH session two criteria; Spearman correlations were assessed for non-parametric comparison given the ordinal scale of some measures (i.e., MTRS and skillfulness items).

CFA of the proposed structure of the AIMS was performed in two phases using MPLUS Version 7.4 (Muthen & Muthen, 1998 – 2014). First, consistent with methods by Carroll and colleagues (2000), model fit by function (i.e., Explore, Teach, Connect) was tested. Model fit was assessed using standard benchmarks including: a non-significant chi square test, root-mean square error of approximation (RMSEA) and standardized root-mean-square residual (SRMR) values of .08 or lower, and a comparative fit index (CFI) of .90 or higher (Bentler, 1990; Hu & Bentler, 1999; Kline, 2011). Second, consistent with methods by Owens and

colleagues (2015), a three factor structure was fit to the data. Here, factors were allowed to correlate and other small modifications were employed to improve model fit. As mentioned above, session two data were selected for primary CFA analyses and reporting.

Results

Inter-rater Reliability

For inter-rater reliability, summary scores for each AIMS function were calculated and both summary level and item level reliability estimates are provided. Table One shows ‘excellent’ (Cicchetti, 1994) reliability for all, but two items. Specifically, ICC values ranged from .783 to .995. The exceptions were Goal Setting (ICC = .610), an explore change item (Explore), and General Information (ICC = .058), a non-function item. These two items were ‘good’ and ‘poor’ respectively (Cicchetti, 1994). For ordinal skillfulness measures, PABAK values were ‘moderate’ for Explore (M = 3.11(SD = .80); k = .415) and Teach (M = 3.09(SD = .89); k = .441) and ‘fair’ for Connect (M = 3.03(SD = .74); k = .362). (Landis & Koch, 1977).

Internal Consistency Reliability

Across treatment sessions, internal consistency analyses showed ‘acceptable’ reliability for the three AIMS summary scores (Nunnally, 1978). For therapist Explore, Cronbach’s alpha was $\alpha = .781$. Therapist Teach and therapist Connect showed alpha values of similar magnitude ($\alpha = .744$ and $\alpha = .788$, respectively).

Factorial Validity

Confirmatory factor models were run by each proposed behavioral treatment function, and standardized regression coefficient loadings are provided in Table Two. Here, coefficient estimates are interpreted as the amount of change in the latent factor when the respective item changes by one unit. These analyses showed generally good model fit, with comparatively better fit for Teach and Connect, in contrast to the Explore function. The Explore function showed the following fit indices: Chi Square = 4.551 ($p = .161$); RMSEA = 0.101 (CI: 0.000–0.226); SRMR = 0.045; CFI = 0.953. The Teach function indices were as follows: Chi Square = 9.687 ($p = .084$); RMSEA = 0.086 (CI: 0.000–0.167); SRMR = 0.050; CFI = 0.957 and the Connect function indices were as follows: Chi Square = 3.644 ($p = .162$); RMSEA = 0.081 (CI: 0.000–0.211); SRMR = 0.036; CFI = 0.950. Next, the full three factor model was fit to the data, and here, indices indicated poor fit. Specific values were as follows: Chi Square = 147.947 ($p < .0001$); RMSEA = 0.123 (CI: 0.119–0.140); SRMR = 0.104; CFI = 0.613. Minor and conceptually appropriate adjustments (e.g., correlated error terms) did not result in substantive improvement. Table Three reports the standardized regression loadings, and similar to the initial CFA results, the Explore function performed the most poorly based on coefficient significance tests. The full correlation matrix of the 15 AIMS items is available upon request from the first author.

Convergent Validity

Convergent validity results for the proposed behavioral treatment functions in relation to existing Project MATCH criterion measures are presented in Tables Four and Five. In these

analyses, both behavior count summaries and skillfulness indicators were examined. Consistent with expectations, the occurrence of therapist exploratory behaviors and therapist skill at Exploring showed positive and significant associations to the criterion MTRS item Depth of Exploration. However, Depth of Exploration also showed positive associations to skillful Teaching and Connecting, and the occurrence of connecting behaviors was more strongly associated with this criterion than exploring behaviors. For therapist Teaching, and consistent with expectations, both occurrence and skillfulness were positively related to the criterion MTRS item Advice Giving. Unexpectedly, therapist connecting behaviors were also positively associated with this criterion. Finally, therapist Connecting was positively and significantly associated with MTRS Empathy, but not Connecting skillfulness.

Discussion

This study presents psychometric findings on a novel observational rating measure of common factors, or *functions*, of behavioral intervention for alcohol use disorder. For research on mechanisms of change in the addictions, this study suggests the Alcohol Intervention Mechanisms Scale is a psychometrically-promising measure of three proposed core processes of treatment – *Exploring and Teaching about change* and *Connecting on an interpersonal level*. Where the measure performs best is with regard to rater reliability, while validity results were mixed. All findings are discussed, and the latter are additionally considered in relation to current process research methodology. Specifically, we are only beginning to understand where classical test theories (i.e., psychometrics) should and should not apply to observations of behavior, and particularly those that may be expected to change over time in response to therapeutic priority, protocol, or clinical need in the moment.

Reliability

The large majority of AIMS behavior count items, and summary scores ranged well above the standard ICC threshold for ‘excellent’ reliability. This illustrates one aspect of the ease of this observational rating measure. In particular, there are relatively few items, and raters can be trained to reliably observe them. The time required for training also compares favorably to other observational process measures in the field. The two lower agreement items, Goal Setting and General Information, were also low occurring items (i.e., roughly two occurrences per session), and this has been found to negatively impact rater reliability (Xu & Lorber, 2014). This result calls into question the necessity of these two items. Internal consistency reliability, however, showed summary function scores including these item were acceptable (Nunnally, 1978). The three function skillfulness indicators were less reliable than behavior count indicators, as determined by PABAK ratings. These indicators represent the ‘competence’ (i.e., adherence and competence; Barber, Krakauer, Calvo, & Badgio, 1997) aspect of the AIMS, and are therefore important. In the present case, ratings clustered at “average” or “good”, which may more substantively suggest revision to verbal anchors (i.e., further explanation) would encourage greater spread in scoring practices. Both of the noted changes will be considered in future iterations of the AIMS, but do not detract from the generally good reliability of the measure.

Validity

Because two people see it, does it make it real? And even if it is real, does it matter? These are questions of reliability versus construct validity and construct versus predictive validity, respectively. While reliability of the AIMS was strong, the validity of the proposed factors was mixed. The single factor models by function were supported by the data, but the three factor model was not. These analyses used session two data, but similar results were found in three other sessions in sensitivity analyses. Poor fit for a multi-factor model was found in recent research on a treatment-specific measure designed to assess Alcohol-focused Behavioral Couples Therapy (i.e., results were on sessions one and eight/nine; Owens et al., 2015). We can speculate reasons for this. There may be barriers to factorial validity that are statistical (e.g., count data with very rare and very frequent items) and/or conceptual (e.g., the processes of therapies are expected to change with person, context, or time). The current study presented an even higher bar since factors were argued to hold up across therapy types rather than one. Yet, even if Exploring, Teaching, and Connecting occur broadly across behavioral treatments, should they occur at similar rates? In a recent study using the AIMS, two broad classes of intervention, characterized by differential reliance on the three common functions, were proposed. Specifically, there are skill-based treatments that rely primarily on teaching and coaching methods and in contrast, there are motivation-based treatments that rely more on an exploratory, client-centered orientation (Magill et al., 2016a). Unfortunately, the scope of the study sample does not support a CFA within condition (i.e., CBT, TSF, MET) or in CBT/TSF (*skill-based*) in contrast to MET (*motivation-based*). In sum, future work would require a larger sample in order to confirm the proposed AIMS factor structure as well as to test factorial invariance by modality and over time.

The AIMS Explore, Teach, and Connect functions, both behavior count and skillfulness measures showed generally good convergent validity with the selected MTRS criterion measures – Depth of Exploration, Advice, and Empathy. The pattern of correlations was consistent with expectations, with the exception of Connect skillfulness, which showed no association to MTRS Empathy. To therapists, the message might be that quantity does not equate quality when it comes to interpersonal connection with clients. Another noteworthy trend in these validity findings is that convergent associations were consistent with expectations, but measures also showed associations with each other. This may shed some light on above difficulties with factorial validity, even if oblimin rotation (i.e., factors were correlated in the converged model) was used. Thus, we conclude that more research is also needed to assess how factors should relate to each other across treatments, within-treatments, and by class of treatments (i.e., *skill-based vs motivation-based*). A final question with respect to validity is - *do these therapeutic functions matter?* Recent work suggests they do, when multiple causal links are considered. Both Explore and Connect behaviors have been shown to sequentially predict subsequent client language about behavior change (Magill et al., 2016a) and client language about changing drinking does predict alcohol outcomes up to one year later (Magill et al., 2016b).

Limitations and Conclusions

This study has some limitations to consider. The use of archival MATCH data can be considered a strength given the availability of three evidence-based treatments for analysis.

This study still suffers the limitations inherent to any secondary analysis design. In our case, the study sample was relatively homogeneous with respect to age, gender, and race. Further, these were aftercare participants who had been exposed to inpatient alcohol treatment, and it is unclear how our results would replicate among MATCH outpatients, or alcohol dependent patients receiving treatment more generally. Given our sample size, we were also unable to test variation in psychometric performance by treatment condition. Finally, we were constrained by the available criterion time points in the MATCH dataset, and this would not have been the case if we had conducted an original study upon which to validate the AIMS measure.

This research is guided by the argument that theories and methods for the study of common factors of behavior change therapies are important and needed. Recent work proposes a classification system to facilitate better design and specification of behavioral therapies. In this model, a Behavior Change Technique is defined as "...an *observable, replicable, and irreducible* component of an intervention designed to alter or redirect causal processes that regulate behavior..." (Michie et al., 2012 p.81). We argue AIMS indicators meet these criteria while allowing for reliable process assessment of alcohol or other drug treatment. Psychometric properties are promising in this preliminary sample, but further work is needed to improve our understanding of its factorial validity. With these refinements, the AIMS may help inform our understanding of how behavioral addictions treatments optimally produce change.

Acknowledgments

AIMS Manual By: Magill, M., & Apodaca, T.R. (2011). Thank you to Cydney Dupree and Colleen Peterson for their thoughtful feedback on the measure. We extend a special thank you to Dr. Theresa Moyers whose research has been influential to the development of this line of study. This project is supported by a Career Development Award awarded to Dr. Molly Magill (K23, AA018126, NIAAA).

References

- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, ... Longabaugh R. Combined pharmacotherapies and behavioral interventions for alcohol dependence: The COMBINE study: A randomized controlled trial. *Journal of the American Medical Association*. 2006; 295(17): 2003–2017. [PubMed: 16670409]
- Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction*. 2009; 104(5):705–715. [PubMed: 19413785]
- Barber JP, Krakauer I, Calvo N, Badgio PC. Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. *Journal of Psychotherapy Practice & Research*. 1997; 6(1):12–24. [PubMed: 9058557]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107(2):238–246. [PubMed: 2320703]
- Carroll KM, Connors GJ, Cooney NL, DiClemente CC, Donovan DM, Kadden RR, Zweben A. Internal validity of project MATCH treatments: Discriminability and integrity. *Journal of Consulting and Clinical Psychology*. 1998; 66(2):290–303. [PubMed: 9583332]
- Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, ... Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*. 2000; 57(3):225–238. [PubMed: 10661673]
- Cicchetti V. Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–90.

- Feinstein AR, Cicchetti DV. High agreement but low kappa: The problems of two paradoxes. *Journal of Clinical Epidemiology*. 1990; 43(6):543–549. [PubMed: 2348207]
- Forcehimes AA, Tonigan JS. Self-efficacy as a factor in abstinence from alcohol/other drug abuse: A meta-analysis. *Alcoholism Treatment Quarterly*. 2008; 26(4):480–489.
- Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*. 2012; 8(1):23–34. [PubMed: 22833776]
- Hofman SG, Barlow DH. Evidence-based psychological interventions and the common factors approach: The beginnings of a rapprochement? *Psychotherapy*. 2014; 51(4):510–513. [PubMed: 25111379]
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6(1):1–55.
- IBM Corporation. *SPSS Statistics (Version 22)*. New York, N.Y: IBM; 2013.
- Imel ZE, Wampold BE, Miller SD, Fleming RR. Distinctions without a difference: direct comparisons of psychotherapies for alcohol use disorders. *Psychology of Addictive Behaviors*. 2008; 22(4): 533–543. [PubMed: 19071978]
- Kadden, R.; Carroll, KM.; Donovan, D.; Cooney, N.; Monti, P.; Abrams, D.; ... Hester, R. *Cognitive-Behavioral Coping Skills Therapy Manual: A clinical research guide for therapists treatment individuals with alcohol abuse and dependence*. Vol. 3. Washington, DC: Government Printing Office; 1992. NIAAA Project MATCH Monograph Series Publication No. (ADM) 92-1895
- Karno M, Longabaugh R, Herbeck D. What explains the relationship between the therapist structure \times patient reactance interaction and drinking outcome? An examination of potential mediators. *Psychology of Addictive Behaviors*. 2010; 24(4):600–07. [PubMed: 21038931]
- Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*. 2007; 3:1–27.
- Kelly JF, Magill M, Stout RL. How do people recover from alcohol dependence? A systematic review of the research on mechanisms of behavior change in Alcoholics Anonymous. *Addiction Research & Theory*. 2009; 17(3):236–259.
- Kiluk BD, Nich C, Babuscio T, Carroll K. Quality versus quantity: Acquisition of coping skills following computerized cognitive-behavioral therapy for substance use disorders. *Addiction*. 2010; 105:2120–2127. [PubMed: 20854334]
- Kline, R. *Principles and Practice of Structural Equation Modeling*. 3. New York, NY: Guilford; 2011.
- Krentzman AR, Cranford JA, Robinson EA. Multiple dimensions of spirituality in recovery: A lagged mediational analysis of Alcoholics Anonymous' principal theoretical mechanism of behavior change. *Substance Abuse*. 2013; 34(1):20–32. [PubMed: 23327501]
- Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977; 33(2):363–74. [PubMed: 884196]
- Laska KM, Gurman AS, Wampold BE. Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy*. 2014; 51:467–481. [PubMed: 24377408]
- Litt MD, Kadden RM, Stephens RS. Coping and self-efficacy in marijuana treatment: Results from the marijuana treatment project. *Journal of Consulting and Clinical Psychology*. 2005; 73(6):1015–2025. [PubMed: 16392975]
- Litt MD, Kadden RM, Cooney NL, Kabela E. Coping skills and treatment outcomes in cognitive-behavioral and interactional group therapy for alcoholism. *Journal of Consulting and Clinical Psychology*. 2003; 71(1):118–128. [PubMed: 12602432]
- Longabaugh R, Magill M. Recent advances in behavioral addiction treatments: Focusing on mechanisms of change. *Current Psychiatry Reports*. 2011; 13(5):382–389. [PubMed: 21750958]
- Longabaugh, R.; Wirtz, PW. *Project MATCH hypotheses: Results and causal chain analyses*. Rockville, MD: NIAAA; 2001.
- Magill, M.; Apodaca, T. *Alcohol Intervention Measurement Scale*. Brown University; 2011. Unpublished manual
- Magill M, Apodaca TR, Karno M, Gaume J, Walthers J, Durst A, ... DiClemente C. Reliability and validity of an observational measure of client decision-making: The Client Language Assessment -

Proximal/Distal (CLA-PD). *Journal of Substance Abuse Treatment*. 2016b; 63(1):10–17. [PubMed: 26898715]

- Magill M, Gaume J, Apodaca TR, Walthers J, Mastroleo N, Borsari B, Longabaugh R. The technical hypothesis of motivational interviewing: A meta-analysis of MI's key causal model. *Journal of Consulting and Clinical Psychology*. 2014; 82(6):973–983. [PubMed: 24841862]
- Magill M, Walthers J, Mastroleo N, Gaume J, Longabaugh R, Apodaca TR. Therapist and client discussions of drinking and coping: a sequential analysis of therapy dialogues in three evidence-based alcohol use disorder treatments. *Addiction*. 2016a; 111(6):1011–1020. [PubMed: 26780564]
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1(1):30–46.
- Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behavior change techniques to reduce excessive alcohol consumption. *Addiction*. 2012; 107(8):1431–40. [PubMed: 22340523]
- Miller, WR.; Moyers, TB.; Ernst, D.; Amrhein, P. Manual for the Motivational Interviewing Skill Code Version 2.0 (MISC). University of New Mexico; 2003. Unpublished manual
- Miller, WR.; Moyers, TB.; Ernst, D.; Amrhein, P. Manual for the Motivational Interviewing Skill Code Version 2.1 (MISC). University of New Mexico; 2008. Unpublished manual
- Miller, W.; Rollnick, S. *Motivational interviewing: Helping people change*. 3. New York, NY: Guilford Press; 2013.
- Miller, WR.; Zweban, A.; DiClemente, CC.; Rychtarik, RG. *Motivational Enhancement Therapy Manual: A clinical research guide for therapists treatment individuals with alcohol abuse and dependence*. Vol. 2. Washington, DC: Government Printing Office; 1992. NIAAA Project MATCH Monograph SeriesPublication No. (ADM) 92-1894
- Morgenstern J, McKay J. Rethinking the paradigms that inform behavioral treatment research for substance use disorders. *Addiction*. 2007; 102(9):1377–89. [PubMed: 17610541]
- Morgenstern J, Longabaugh R. Cognitive-behavioral treatment for alcohol dependence: A review of evidence for its hypothesized mechanisms of action. *Addiction*. 2000; 95(10):1475–1490. [PubMed: 11070524]
- Moyers TB, Martin TB, Christopher PJ, Houck JM, Tonigan JS, Amrhein PC. Client language as a mediator of motivational interviewing efficacy: Where is the evidence? *Alcoholism: Clinical & Experimental Research*. 2007; 31(Suppl S3):40s–47s.
- Muthen, & Muthen. *MPLUS (Version 7.4)*. Los Angeles, CA: Muthen & Muthen; 2014.
- Nowinski, J.; Baker, S.; Carroll, KM. *Twelve-Step Facilitation Therapy Manual: A clinical research guide for therapists treatment individuals with alcohol abuse and dependence*. Vol. 1. Washington, DC: Government Printing Office; 1992. NIAAA Project MATCH Monograph SeriesPublication No. (ADM) 92-1893
- Nunnally, JC. *Psychometric Theory*. 2. New York, NY: McGraw-Hill; 1978.
- Nuro, KF.; Maccarelli, L.; Martino, S.; Ball, SA.; Baker, SM.; Rounsaville, BJ.; Carroll, KM. *Yale Adherence and Competence Scale (YACSII) Guidelines*. West Haven, CT: Yale University; 2005.
- Owens MD, McCrady BS, Borders AZ, Brovko JM, Pearson MR. Psychometric properties of the System for Coding Couples' Interactions in Therapy-Alcohol. *Psychology of Addictive Behaviors*. 2014; 28(4):1077–1088. [PubMed: 25528049]
- Petry NM, Litt MD, Kadden RM, Ledgerwood D. Do coping skills mediate the relationship between cognitive-behavioral therapy and reductions in gambling in pathological gamblers? *Addiction*. 2007; 102:1280–1291. [PubMed: 17624978]
- Prochaska, JO.; DiClemente, CC. Toward a comprehensive model of change. In: Miller, WR.; Heather, N., editors. *Treating Addictive Behaviors*. New York, NY: Plenum Press; 1986. p. 3-27.
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH Posttreatment drinking outcomes. *Journal of Studies on Alcohol*. 1997; 58(1):7–29. [PubMed: 8979210]
- Project MATCH Research Group. Project MATCH secondary a priori hypotheses. *Addiction*. 1998; 92(12):1671–1698.

- Tober G, Clyne W, Finnegan O, Farrin A, Russel I. Validation of a scale for rating the delivery of psycho-social treatments for alcohol dependence and misuse: The UKATT process rating scale (PRS). *Alcohol and Alcoholism*. 2008; 43(6):675–682. [PubMed: 18818190]
- UKATT Research Team. UK Alcohol Treatment Trial: client-treatment matching effects. *Addiction*. 2008; 130(2):228–238.
- Wampold, BE. *The great psychotherapy debate: Model, methods and findings*. Mahwah, NJ: Lawrence Erlbaum; 2001.
- Wampold BE, Mondin G, Moody M, Stich F, Benson K, Ahn H. A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, 'all must have prizes'. *Psychological Bulletin*. 1997 Nov; 122(3):203–15.
- Xu S, Lorber MF. Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology*. 2014; 82(6):1219–1227. [PubMed: 25090041]

HIGHLIGHTS

- The Alcohol Intervention Mechanisms Scale (AIMS) measures therapist interventions that occur broadly across modalities of behavioral treatment for alcohol use disorder.
- Interrater agreement reliability for the AIMS was “excellent”.
- Internal consistency reliability for the AIMS was “acceptable”.
- In a two stage confirmatory factor model, fit by function was good, but the three factor solution did not fit the data well.
- In concurrent validity analyses, AIMS summary and skillfulness scores showed associations with relevant Project MATCH criterion measures that were consistent with expectations.
- In this preliminary sample, the AIMS shows generally promising psychometric properties.

Table 1

Reliability and Session-level Descriptive Information

AIMS Behavior code	ICC ^a	Minimum ^b	Maximum ^b	Mean ^b	(SD) ^b
Function: Explore					
Explore Change: Question	.99	1	219	67.95	38.12
Explore Change: Reflection	.99	1	251	50.83	33.30
General Assessment	.88	0	121	9.50	15.58
Goal Setting	.61	0	16	1.87	2.57
Explore Summary	.99	8	398	130.15	62.79
Function: Teach					
Teach/Advise	.97	0	167	46.28	28.09
Homework: Teach	.86	0	31	3.91	5.05
Homework: Explore	.80	0	29	2.60	4.17
Self-Disclose	.94	0	64	5.64	8.07
Structure	.97	1	71	20.72	12.99
Teach Summary	.98	10	214	79.13	39.95
Function: Connect					
Affirm	.94	0	57	15.11	10.66
Express Empathy/Support	.94	0	25	3.51	3.85
Emphasize Control/Collaboration	.90	0	25	4.28	3.75
Connect Summary	.94	0	68	22.90	13.90
Other Therapist Behaviors					
Confront	.78	0	15	1.22	2.79
Information	.01	0	15	0.84	1.85
Follow/Neutral/Facilitate	.99	3	212	46.40	32.15

Notes.

^aReliability estimates based on $N = 47$ sessions double-coded. Cicchetti (1994) suggests the following guidelines for assessing reliability of observational coding systems: ICC of .75 or above = excellent; .60-.74 = good; .40-.59 = fair; below .40 = poor. Fair or poor items are shown in **bold**.

^bSession-level descriptive data based on $N = 484$ sessions

Table 2

Confirmatory Factor Analysis by function

AIMS Behavior code	coef ^a	SE ^a
Function: Explore		
Explore Change: Question	.55	.12***
Explore Change: Reflection	.97	.19***
General Assessment	-.16	.09 [†]
Goal Setting	.28	.11**
Function: Teach		
Teach/Advise	1.03	.09***
Homework: Teach	.63	.07***
Homework: Explore	.44	.08***
Self-Disclose	.04	.09
Structure	.30	.08***
Function: Connect		
Affirm	.47	.12***
Express Empathy/Support	.55	.13***
Emphasize Control/Collaboration	.49	.12***

Notes. Standardized coefficients.

^aData reported are from session two.

[†] $p < .10$,

** $p < .05$,

*** $p < .0001$.

Table 3

Confirmatory Factor Analysis – three function model

AIMS Behavior code	coef ^a	SE ^a
Function: Explore		
Explore Change: Question	.35	.19 [†]
Explore Change: Reflection	1.58	.84 [†]
General Assessment	-.10	.08
Goal Setting	.09	.12
Function: Teach		
Teach/Advise	.93	.07 ^{***}
Homework: Teach	.68	.06 ^{***}
Homework: Explore	.33	.09 ^{***}
Self-Disclose	.02	.09
Structure	.49	.08 ^{***}
Function: Connect		
Affirm	.47	.12 ^{***}
Express Empathy/Support	.55	.13 ^{***}
Emphasize Control/Collaboration	.49	.12 ^{***}
Explore WITH Teach	-.24	.14
Explore WITH Connect	.18	.15
Connect WITH Teach	-.08	.14

Notes. Standardized coefficients.

^aData reported are from session two. Session two data were re-run, dropping non-significant items and assessing a two factor structure (Teach; Connect); this resulted in greatly improved, but unacceptable model fit. The pattern of results was consistent when CFA were run at other available timepoints (i.e., first, third, final).

[†] $p < .10$,

** $p < .05$,

*** $p < .0001$.

Table 4

AIMS Therapist Function Summary Scores

	AIMS functions			Selected MTRS items		
	Sum of Explore	Sum of Teach	Sum of Connect	Depth of Exploration	Advice Giving	Empathy
AIMS functions						
Sum of Explore	1	.016	.265***	.184*	-.099	.046
Sum of Teach		1	.279***	.157 [†]	.330***	-.081
Sum of Connect			1	.298***	.341***	.211*
Selected MTRS items						
Depth of Exploration				1	.253***	.361***
Advice Giving					1	-.059
Empathy						1

Notes.

^aData reported are from session two. *N* = 121.

[†]<.10,

* <.05,

** <.01,

*** .005

Table 5

AIMS Therapist Function Skillfulness Scores.

	AIMS functions			Selected MTRS items		
	Therapists Skill at Explore	Therapists Skill at Teach	Therapists Skill at Connect	Depth of Exploration	Advice Giving	Empathy
AIMS functions						
Therapists Skill at Explore	1	.500***	.585***	.416***	.014	.178*
Therapists Skill at Teach		1	.471***	.252***	.212*	.086
Therapists Skill at Connect			1	.256***	.158 [‡]	.090
Selected MTRS items						
Depth of Exploration				1	.253***	.361***
Advice Giving					1	-.059
Empathy						1

Notes.

^aData reported are from session two. *N* = 121.

[‡]<.10,

* .05,

** <.01,

*** .005.