Published in final edited form as:

J Subst Abuse Treat. 2017 December ; 83: 55–61. doi:10.1016/j.jsat.2017.10.005.

Measuring Sustainment of an Evidence Based Treatment for Adolescent Substance Use

Wenjing Huang, Ph.D.¹, Sarah B. Hunter, Ph.D.¹, Lynsay Ayer, Ph.D.², Bing Han, Ph.D.¹, Mary E. Slaughter, M.S.³, Bryan R. Garner, Ph.D.⁴, and Susan H. Godley, Rh.D.⁵

¹RAND, Santa Monica, CA 90407-2138

²RAND, Arlington, VA 22202-5050

³RAND, Pittsburgh, PA 15213

⁴Research Triangle Institute, Raleigh, NC 27675

⁵Chestnut Health Systems, Normal, IL 61761

Abstract

Background—Measuring the extent a behavioral treatment is sustained in usual care practice settings after initial support for implementation ends is challenging. Oftentimes dichotomous self-report measures are used to measure health program sustainment, but these measures may not adequately capture the quality or extent of behavioral treatment delivery. Using data collected from community based organizations that received support to implement the Adolescent Community Reinforcement Approach (A-CRA), a measure of the extent of sustainment was derived.

Methods—Based on implementation theories and implementation support protocols, a total of 10 core treatment elements were identified to measure the extent of sustainment using information collected from key clinical staff. Item response theory (IRT) and principal component analyses (PCA) were used to further refine the 10 elements into composite measures of sustainment. The association between the 10 elements and the relationships between the comprehensive measures to a dichotomous sustainment measure were also examined.

Results—Results from PCA identified two components from the initial 10 elements for measuring the extent of A-CRA treatment sustainment. The two components described different aspects of organizational support for A-CRA treatment implementation, one representing the quality or extent of treatment staffing and delivery and the other representing the quality of or extent of clinical supervision.

Conclusions—Using IRT and PCA, we were able to derive components that could be used to measure the extent of EBT sustainment and also better capture the quality of treatment delivery

Corresponding Author: Wenjing Huang, Ph.D., RAND, Santa Monica, CA 90407-2138, whuang@rand.org.

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.

than the use of simple dichotomous measure. The methodological contribution of our study is that we have demonstrated a general analytic approach that may be applicable for other psychosocial treatments.

Keywords

sustainment; evidence-based treatments; methodology

1. Introduction

While the use of efficacious interventions is critical for improving treatment quality, they hold little value without attention to the quality of their implementation and sustainment (Simpson & Flynn, 2007). Sustainment is defined as the continued use of an innovation in practice following initial adoption and implementation (Aarons, Hurlburt, & Horwitz, 2011). The sustainment of evidence based treatment (EBT) is a major concern of many funders, researchers, practitioners and policy makers, as many evidence-based behavioral interventions fail to reach their full potential when delivered in real world practice settings (Aarons, Hurlburt & Horwitz, 2011). For instance, an EBT may be implemented in the context of a grant-funded research or implementation support effort, but not *sustained* following the end of funding. Thus, initial investments often do not achieve the maximum public health impact possible (Aarons et al., 2014; Hodge, Turner, Sanders & Filus, 2016).

After initial funding for the implementation or testing of an EBT ends, resources are seldom available to continue studying whether it is sustained (Scheirer, Hartling & Hagerman, 2008). Moreover, recent systematic reviews note that it has been a challenge to measure EBT sustainment (Wiltsey Stirman, Kimberly, Cook, Calloway, Castro & Charns (2012); Ament et al., 2015). Measuring the sustainment of a behavioral treatment is particularly complex because unlike other types of treatments or public health programs that have been studied, it typically involves multiple therapeutic sessions covering a range of topics that can be adapted or tailored to individual needs.

The quality or extent of EBT implementation and its continued use (i.e., sustainment) has been conceptualized to include several elements including the organizational capacity to ensure treatment delivery is maintained at the program or organizational level (e.g., adequate clinical supervision and training; Fixsen, Naoom, Blasé, Friedman & Wallace, 2005). Operationalizing EBT sustainment in this way is consistent with recommendations from a recent systematic review (Wiltsey Stirman et al., 2012) and was recently employed by Aarons and colleagues (Aarons et al., 2016; Green et al., 2015). More specifically, Wiltsey Stirman et al (2012) recommended that the operationalization of program sustainment include the capacity elements needed to function at the required level to maintain program benefit. In the Aarons et al (2016) study, this was further clarified as an assessment of “structural fidelity”, that is, elements such as adequate supervision (e.g., coaching visits, regularly scheduled team meetings) to support program delivery.

Despite recent these developments in understanding EBT sustainment, the assessment of health program or treatment sustainment has often been based on simple dichotomous staff reports as to whether it is offered or not (Scheirer & Dearing, 2011). However, a self-

reported dichotomous indicator may have a high false positive rate (Adams et al., 1999; Ament et al., 2015), and does little to assess the many provider and organizational characteristics that reflect the extent or quality of behavioral treatment delivery. Apart from self-reports, organizational assessments that include chart reviews, interviews and site observations by trained experts have also been used to measure the extent or quality of health program implementation or sustainment (e.g., Dual Diagnosis Capability in Addiction Treatment; McGovern, Matzkin, & Giard, 2007). However, the cost of having well-trained researchers to conduct such assessments that require in-person site visits can be high. As noted in the emerging field of implementation science, more feasible, rigorous methods and measures are needed in order to assess EBT sustainment (Lewis et al., 2015; Proctor et al., 2015). A sustainment measure based on a more comprehensive assessment of the varied organizational and individual provider level components of sustainment is thus needed to better capture the quality of or extent to which a behavioral treatment is sustained.

Therefore, we set out to develop a comprehensive and refined assessment of EBT sustainment that could be utilized without the requirement of in-person site visits by trained professionals. We developed this assessment within the context of a study examining the sustainment of the Adolescent-Community Reinforcement Approach (A-CRA), an EBT for problematic adolescent substance use. A-CRA is a behavioral intervention that encourages adolescents to engage in pro-social activities and behaviors that promote recovery and minimize the impact of environmental contingencies that support substance use. Five randomized controlled studies of A-CRA have shown positive effects on alcohol use, mental health, and social domains (Dennis et al., 2004; Godley et al., 2007; Slesnick, Prestopnik, Meyers & Glassman, 2007; Godley et al., 2014; Henderson et al., 2016). From 2006–2013, the Substance Abuse and Mental Health Services Administration provided implementation support to community-based treatment organizations located in 27 states, districts or territories across the United States. We examined the extent of A-CRA sustainment in these organizations following the end of the SAMHSA grant funding.

The aim of this research was to develop a measure of the extent of A-CRA sustainment that incorporated both provider-level and organizational-level elements. Based on a review of implementation literature and the implementation support activities monitored during the funding period, we identified several provider- and organizational-level EBT sustainment elements to assess in this study. These elements were used to derive a comprehensive measure of EBT sustainment. To assess its external validity, we compared our new measure of sustainment to a dichotomous self-report sustainment measure (i.e., yes/no) that is more commonly used to assess sustainment of health programs more broadly.

2. Material and Methods

2.1. Participants

Eighty two of 84 funded organizations were no longer receiving federal funding to implement A-CRA at the time of data collection (Fall 2013) and were thus targeted for the study. Two organizations were in a no-cost extension period and still using grant support to deliver A-CRA. We recruited both clinicians providing direct services to youth and clinical

supervisors of those clinicians. We aimed to recruit at least one clinician and one clinical supervisor from each organization.

2.2. Data sources

Phone interviews, web-based surveys, and training documentation were collected from staff to assess the extent of A-CRA sustainment.

2.3. Procedures

Our research team obtained the contact information of staff at the funded treatment organizations from the organization that provided A-CRA implementation support during the SAMHSA funding period. Via mail, phone and email we invited currently employed clinicians and clinical supervisors at the 82 organizations to participate in the study. All participants provided informed consent prior to enrolling in the study. Semi-structured interviews were conducted by trained survey field staff. All interviews were digitally recorded and field notes were transcribed. The transcribed notes were then reviewed by a second research assistant for accuracy. Once an interview was completed, participants were invited to complete a web-based survey. Incentives for the completion of interviews and web surveys were provided.

2.4. Measures

2.4.1. Sustainment elements—We assessed 10 sustainment “elements” a priori that were thought to be necessary for quality A-CRA implementation. We identified these elements by first consulting the previous implementation literature (e.g., Fixsen et al., 2005; Scheirer and Dearing, 2011) and then enhanced the list based on input from the A-CRA implementation support team and developer on what aspects of the clinical treatment protocol were monitored during the funding period (see Godley et al., 2011). For example, from the literature we identified that “dosage” was considered a key element, and for A-CRA, we applied the specific clinical treatment guidelines of 11–15 sessions to assess this element. In sum, the 10 elements represent key provider and organizational characteristics that are needed to support quality implementation of A-CRA. The operational definitions of each of these characteristics and how the data were collected are described next.

1) Clinical knowledge of A-CRA: This element was based on responses to 10 multiple choice web survey questions that were derived from the A-CRA treatment manual which covered the theoretical underpinnings and the content of selected A-CRA procedures (Godley et. al., 2001). Each item included five response options and answers were coded as correct (scored as a “1”) or incorrect (scored as a “0”) for a total score range from 0–10. During the funded implementation phase, clinicians and supervisors were required to demonstrate proficiency on these items in order to obtain A-CRA certification. Multilevel Item Response Theory (MIRT) was used to create a more refined measurement of clinical knowledge (see more details in the following MIRT subsection).

2) Presence of certified A-CRA clinician(s): This element was assessed through asking clinical supervisors whether their organization currently had a clinician on staff in their adolescent substance use treatment program that was certified to deliver A-CRA, and were

asked to provide the name(s) of any certified staff. Responses from participants were verified using the treatment developer records that documented who had successfully completed the certification process.

3) Presence of certified A-CRA supervisor(s): This element assessed whether an organization currently employed a staff person certified to provide A-CRA clinical supervision using similar methods as described for clinician.

4) Clinical supervisors' knowledge of A-CRA training and certification process: This element assessed the clinical supervisors' knowledge about the A-CRA clinician training and certification process based on the percentage of correctly answered 15 true/false questions. These statements were drawn from a list of requirements of the A-CRA certification process and outlined in the A-CRA coding manual (Smith, Lundy, & Gianini, 2007) that supervisors were required to master as part of the A-CRA treatment certification process during the implementation support period. In organizations with more than clinical supervisor, the percentage correct was averaged to create an organization level value that ranged from 0–100%.

5) Percentage of youth receiving A-CRA: Clinical supervisors were first asked in the interview about how many youths had received substance use treatment at their organization in the past six months. Then they were asked how many youths received A-CRA during that same period. A percentage of youth clients receiving A-CRA (of those receiving any substance use treatment) was then calculated for each organization (range 0 – 100%).

6) A-CRA dosage: This element was assessed by asking clinicians who delivered A-CRA an interview question about how many A-CRA treatment sessions they told clients they would receive. If all respondents at an organization endorsed within the treatment developers' recommended range of sessions (i.e., 11–15 sessions), the organization received a score of “1”, otherwise the organization scored “0” on this element.

7) Clinical supervision frequency: This element was assessed by asking clinicians in the interview how often they received clinical supervision. Bi-weekly clinical supervision is emphasized as part of the A-CRA training protocol (Godley et al., 2011). Organizations where clinicians reported receiving clinical supervision every two weeks or more were coded a “1” and organizations where clinicians reported receiving clinical supervision less frequently were coded a “0”. If more than one clinician responded per organization, the responses were averaged for an organization-level score.

8) Clinical supervision content: This element was designed to measure the quality of clinical supervision. Clinicians were asked to respond to six interview questions assessing specific supervision activities. The items were based on the rating manual used for coding recorded supervision sessions during the supervision certification process (Hupp, Mertig, Krall, Godley & Godley, 2009). An example item used was, “My supervisor reviews my A-CRA case review report” with “yes” and “no” as response options. Responses to the six items were scored correct (coded as “1”) or incorrect (coded as “0”).

9) Clinical supervision with treatment session review: This element assessed whether clinicians reported that their clinical supervisor reviewed treatment sessions and a) told the clinician what they had done well and/or b) gave suggestions to the clinician about how they could improve treatment delivery. These items were based on guidelines in the A-CRA treatment manual. Responses were scored “0” if clinicians reported that their clinical supervisors did not review sessions; “0.5” if the sessions were reviewed and clinicians were told they had done well; or “1” if session review also included suggestions on how treatment delivery could be improved. If more than one clinician responded per organization, the responses were averaged for an organization-level score.

10) Onsite A-CRA clinical training quality: This element was assessed by evaluating the content of the A-CRA clinical training provided onsite. Clinical supervisors who reported providing A-CRA training at their organization were asked to submit a training agenda. Submission of a training agenda was one of the required elements for the A-CRA Clinical Supervisor Certification process and supervisors were provided feedback until their agendas met specified criteria. For this study, submitted agendas were rated by a member of the A-CRA implementation support team for its coverage of key A-CRA training aspects. Scores ranged from 0 – 1 representing the proportion of A-CRA training aspects covered.

2.4.2. A-CRA sustainment status—An indicator of A-CRA sustainment was collected by asking participants in the interview if A-CRA was still being offered to youth at their organization and created a dichotomous indicator (i.e., “0” indicated no longer offering A-CRA and “1” indicated still offering A-CRA) that then could be examined in relation to the other measures. This item was asked at the beginning of the interview, prior to any other questions about the A-CRA treatment elements (as detailed in 2.4.1).

2.5. Analytic Plan

2.5.1. Data pre-processing—Elements 1 (clinical knowledge of A-CRA) and 8 (clinical supervision content) were based on multiple items where we could assume the existence of underlying latent constructs, whereas the other elements consisted of responses around a single domain. Thus, in order to prepare the data regarding these elements for the main analyses, we used item response theory (IRT; see Embretson & Reise, 2000). IRT models are suitable for analysis of questionnaire or test response data. The multiple items of these two elements were summarized in a latent variable modeling framework so that measurement errors were taken into account. In addition, each item contributed differently to measuring the latent construct representing the corresponding element.

Since there were frequently multiple respondents (varying from 1–3) within the same organization that responded to these items, the hierarchical structure of the data required using Multilevel Item Response Theory (MIRT; Reckase, 2009). We used a 2-level Rasch model (Rasch, 1960) implemented in the computer software flexMIRT (Cai, 2012) to obtain scoring of the underlying latent construct at both the individual- and the organization-level. To be consistent with the other elements and the main purpose of the study (i.e., to create an organization-level measure), we used only the organization-level IRT scores as the final element scores.

Since PCA requires complete data, we addressed the issue of missing data through multiple imputation (Rubin, 1987) using SAS PROC MI (SAS Institute Inc., 2015). Twenty data sets were imputed. We used a SAS macro provided by Cai and Lee (2009) to first calculate the covariance matrix from each of the imputed data set and then to combine the 20 matrices into a single covariance matrix.

2.5.2. Principal Component Analysis (PCA)—Next, the relation of how the 10 elements were related to one another was examined with the goal to develop a more comprehensive and refined assessment of the extent of A-CRA sustainment. We conducted a standard PCA analysis for the covariance matrix of the 10 elements using SAS PROC FACTOR.

2.5.3. Comparing derived sustainment components to dichotomous sustainment indicator—To investigate how each derived component from the PCA was related to a dichotomous indicator of sustainment status, a series of logistic regressions were fitted, testing the effect of each component derived from the PCA with the dichotomous sustainment indicator as the outcome.

3. Results

3.1. Sample Characteristics

One hundred and twenty respondents (i.e., 53 clinicians, 34 clinical supervisors, and 33 individuals that reported they served as both clinicians and clinical supervisors at their organization) from 68 out of 82 organizations provided data (representing an 83% response rate). Staff from 46 organizations reported that their organization was continuing to deliver A-CRA (i.e., 67% of the sample). Time since funding loss ranged from 1 to 48 months. Response rates were not related to time since funding loss as we found no significant relation between funding year and participation rate ($\chi^2(4, N = 68) = 3.53, p = 0.53$).

3.2. Preliminary Data Analyses

We combined data from multiple respondents to obtain an organization-level score for each of the 10 elements. The site-level response rate on the different elements ranged from 59% (Element 9: Supervision that includes review of treatment sessions) to 100% (Elements 4: A-CRA training and certification process knowledge). Missing data were handled via multiple imputation and a single covariance matrix was derived based on 20 sets of imputed data.

3.2.1. Using MIRT to Assess Item Characteristics—Results from the 2-Level Rasch MIRT models conducted with Elements 1 and 8 included location parameter estimates that depict item characteristics (e.g., difficulty). These estimates were used to calculate IRT scores. Using Element 8 (i.e., clinical supervision content knowledge) as an example, six items about different supervision activities were assessed for their relative level of importance for measuring the underlying latent construct of clinical supervision content. Factor scores were computed using the software package flexMIRT.

Table 1 presents descriptive statistics for the 10 elements. These 10 organization-level element scores were used as input data for the PCA analyses.

3.2.2. Principal Component Analysis—For the PCA, we used the organization-level MIRT derived scores for elements 1 and 8. For all other elements we used the 0–1 scales previously described. Orthogonal rotation was used in the analyses. Results of initial PCA suggested a three-component structure among the 10 elements according to eigenvalues (3 eigenvalues >1: 2.93, 1.89 and 1.14 respectively). However, the first two components were positively loaded on distinct sets of elements, while the third component was cross loaded on elements overlapping with the first two components. To keep the results more interpretable, we thus selected two components where the first six elements loaded higher on the first component and the next four elements (i.e., 7–10) loaded higher on the second component. See Table 2 for the detailed loading coefficients by elements for the two principal components where the higher loadings are shaded.

The two principal components that were derived have a relatively clear interpretation. Specifically, the first principal component positively loaded on elements 1 to 6, which could be interpreted as summarizing overall organizational capacity regarding treatment staffing and treatment delivery. That is, the component assessment whether the organization had adequate staff and delivery capacity. For example, elements 2 and 3 (i.e., presence of a certified clinician and a certified supervisor) were highly indicative of the staffing aspect with estimated loadings of 0.8 and 0.75 respectively. Elements 5 (i.e., the proportion of eligible youth receiving A-CRA) and 6 (i.e., A-CRA dosage) help describe the extent of treatment delivery.

In contrast, principal component #2 positively loaded on elements 7 to 10 which can be interpreted as assessing the clinical supervision component of A-CRA sustainment with loadings ranging between 0.33 and 0.71. The frequency of supervision (i.e., element 7, loading = 0.58) and the supervision content (i.e., element 8, loading = 0.66) are both important elements to indicate good supervision practice. Session review (i.e., element 9 with estimated loading=0.71) was found to be the most important among the four elements of supervision. The training process (i.e., element 10, a training process that meets quality requirements with estimated loading of 0.33) was found to be relatively less important than the other three elements.

There is a stronger negative loading of element 1 (clinical knowledge) on component #2 (loading = −0.51) than on component #1 (loading = 0.35). The negative loading suggests that those with higher levels of clinical knowledge of A-CRA report lower intensity/quality of clinical supervision, which may make sense in community agencies, especially given that 56% of respondents were either clinical supervisors or both clinicians and clinical supervisors. Those with advanced skills may receive less supervision. Similarly, there is a low negative loading of element 6 (A-CRA dosage) on component #2 (loading = −0.39), which also make sense because those that deliver the recommended range of A-CRA sessions may require less supervision. Element 10 (training quality) cross-loads on both components, suggesting that it contributes to measurement of both dimensions despite its slightly higher loading on supervision than on staffing and delivery. Since we used orthogonal rotation to keep the original orthogonality of the principal components as well as simplicity of interpretation, correlations among the underlying dimensions might manifest themselves as cross-loadings as shown with items 6 and 10.

3.2.3. Comparing Derived Components to Dichotomous Sustainment Indicator

—Next, we calculated composite scores for the two PCA components 1) treatment staffing and delivery and 2) clinical supervision. To do this, we first standardized all the elements; then, we calculated the composite scores as a weighted sum of the standardized scores of each element where weights were the standardized scoring coefficients from PCA (see Table 2). Next, we conducted a series of logistic regressions using the two composite scores to examine their relationship to the dichotomous sustainment indicator. Logistic regression results showed that the treatment staffing and delivery component was statistically significantly related to the dichotomous sustainment indicator ($t = -6.59, p < 0.05$) whereas the supervision component was not ($t = 0.82, p > 0.05$).

Figures 1 and 2 provide visual illustrations of the distributions of the composite scores for components #1 and #2 by the dichotomous sustainment status indicator. Figure 1 shows organizations where staff reported sustaining A-CRA also exhibited higher levels of organizational support in terms of treatment staffing and delivery than those organizations where staff reported not sustaining A-CRA (i.e., sustaining organization $M = 4.77$; $SD = 1.54$; non-sustaining organization $M = 1.61$; $SD = 2.39$; $p < 0.05$). However, when comparing the scores for component #2 (see Figure 2), the clinical supervision levels between the organizations where staff reported sustaining A-CRA and those reported not sustaining did not significantly differ.

4. Discussion

This study focused on understanding how to best conceptualize and measure the extent of sustainment of one of the most widely disseminated EBTs for adolescent substance use, the Adolescent-Community Reinforcement Approach (A-CRA). In this study, different measures that characterized the extent to which the A-CRA treatment model was sustained were examined using data collected from staff from community based adolescent treatment programs following an initial three-year implementation support period. Two main findings were found. First, we successfully empirically derived two meaningful composite assessments of the extent of A-CRA sustainment, one that comprised of treatment staffing and delivery elements and the other comprised of clinical supervision elements. Second, we examined how well these composite measures of A-CRA sustainment were related to a more commonly used and simple sustainment measure, that is, a dichotomous self-report indicator of whether staff reported delivering A-CRA. The treatment staffing and delivery component was significantly related to the dichotomous sustainment indicator, but the supervision component was not. This suggests that the dichotomous sustainment measure is better aligned with the treatment staffing and delivery aspects but does not align well with clinical supervision quality. Empirical evidence has shown that quality clinical supervision is critical for quality implementation of evidence based treatment practices (Miller et al., 2004; Martino, 2010; Schoenwald et al., 2009). Thus, the composite measures of sustainment developed in the current study allow for more detailed understanding of the quality of A-CRA sustainment than a more simple dichotomous indicator. A key conclusion of this study is that meaningful composite measures of A-CRA sustainment that include the assessment of the extent of treatment staff, treatment delivery and clinical supervision that is aligned

with the EBT implementation model may be a more comprehensive and valid measure of EBT sustainment than a simple dichotomous indicator.

Based on the findings in this study, the presence of qualified staff that were knowledgeable about A-CRA and a treatment process closely aligned with A-CRA guidelines appeared related to staff reports of continuing A-CRA. The quality of clinical supervision offered at the organizations was less aligned with the self-reported A-CRA sustainment indicator. This suggests that the clinical supervision component may or may not be sustained at the same level as the other treatment staffing and treatment delivery components when staff consider whether their organization sustains A-CRA. Of note, three of the four clinical supervision elements measured in this study were not specific to A-CRA (i.e., frequency of supervision and content elements) but could exist in regard to any treatment program. However, in fact, most staff reported very low levels on the clinical supervision elements regardless of sustainment status, suggesting that this component was the least likely to be sustained at the levels recommended and supported during the grant-funded implementation support period.

These findings suggest that when an organization loses grant support for implementation, the organization may continue to deliver the EBT as long as the organization continues to employ adequately trained staff. However, quality clinical supervision is less likely to be sustained. This is important because ongoing clinical supervision, including the capacity to train and certify staff in the EBT onsite, is likely going to enhance the quality of EBT delivery and provide the support necessary to prevent staff turnover and drift from the treatment model (Miller et al., 2004; Martino, 2010; Schoenwald et al., 2009). It is also likely though that clinical supervision support is harder to maintain when federal funding ends because there may be less flexible funding to support it.

Though applied specifically to A-CRA in this study, the approach used to identifying elements and developing composite measures of treatment sustainment can be extended to more general scenarios for measuring the extent of EBT implementation or sustainment. This is a unique methodological contribution as IRT and PCA approaches are rarely used in implementation science. These same methodological approaches could be applied to develop measures of sustainment of other EBTs at any point in time or over time to capture the extent of treatment adoption and implementation by an organization.

4.1. Limitations

There are a few important limitations to this study. First, the analyses were based on a relatively small dataset of 68 organizations. Although 68 organizations across multiple U.S. states receiving the same training and support for one EBT is a large sample from a clinical and implementation science perspective, this sample size is still considered small from a statistical point of view. The relative precision and certainty of our findings would be improved with a larger sample. Second, relatively substantial proportions of data were missing on some elements. For example, staff from the majority of organizations had not provided onsite training or certification, therefore no training agendas or site certification information were available for analyses. However, this missing data finding is notable because it suggests a longer-term challenge with sustaining A-CRA even in those organizations sustaining at the measurement points for this study. Inevitably there will be

turnover in staff certified in the intervention and if the organization does not have the ability to train new hires in the intervention, it will not be sustained. Third, we only interviewed data from 1–3 staff at each organization. While some organizations had a larger group of staff serving adolescents, we found that the median number of clinical supervisors and clinicians at the participating sites was 2. Resources limited us from a full query of all employed staff, we focused our efforts on targeting individuals who were primarily employed to serve youth and considered the “lead” at their organization in regards to A-CRA implementation (e.g., a full-time staff member over a part-time staff member). Fourth, our study relied mainly on self-reported data. In some cases, we were able to verify the data using a secondary source (e.g., certification documentation), but not all measures were subject to validation and future research is warranted to focus on this endeavor. Lastly, although the data collection was conducted roughly during the same calendar months, the measurements were still unsynchronized with respect to the implementation period, because each site has a different support period. Unfortunately, we were not able to adjust for the potential temporal trend in our PCA.

4.2. Conclusions

In sum, the general approach used in this study to derive EBT sustainment composite measures may benefit a much broader audience in implementation science that is interested in measuring the quality of implementation or sustainment of similar behavioral interventions that require several dimensions such as qualified staff, specific dosages of treatment and quality clinical supervision. These findings suggest that it is important to assess multiple aspects of EBT sustainment and that it is feasible to do so with relatively straightforward psychometric approaches. Future research using this approach with other EBTs and larger samples is needed to determine its generalizability. Using comprehensive measures of EBT sustainment like the one tested in this study, future research should also examine how EBT sustainment is associated with patient outcomes over time.

References

1. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*. 2011; 38(4): 4–23. [PubMed: 21197565]
2. Aarons GA, Green AE, Trott E, Willging CE, Torres EM, Ehrhart MG, Roesch SC. The Roles of System and Organizational Leadership in System-Wide Evidence-Based Intervention Sustainment: A Mixed-Method Study. *Administration and Policy in Mental Health*. 2016; 43(6):991–1008. [PubMed: 27439504]
3. Adams AS, Soumerai SB, Lomas J, Ross-Degnan D. Evidence of self-report bias in assessing adherence to guidelines. *International Journal for Quality in Health Care*. 1999; 11(3):187–192. [PubMed: 10435838]
4. Ament SM, de Groot JJ, Maessen JM, Dirksen CD, van der Weijden T, Kleijnen J. Sustainability of professionals' adherence to clinical practice guidelines in medical care: a systematic review. *BMJ Open*. 2015; 5(12):e008073.doi: 10.1136/bmjopen-2015-008073
5. Cai, L. flexMIRT®: Flexible multilevel item factor analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group, LLC; 2012.
6. Cai L, Lee T. Covariance structure model fit testing under missing data: an application of the supplemented EM algorithm. *Multivariate Behavioral Research*. 2009; 44(2):281–304. DOI: 10.1080/00273170902794255 [PubMed: 26754269]

7. Dennis M, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J, et al. The Cannabis Youth Treatment (CYT) study: Main findings from two randomized trials. *Journal of Substance Abuse Treatment*. 2004; 27(3):197–213. [PubMed: 15501373]
8. Embretson, SE., Reise, SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
9. Fixsen, D., Naoom, S., Blasé, K., Friedman, R., Wallace, F. Implementation Research: A Synthesis of the Literature. Tampa, FL: The National Implementation Research Network, Louis de la Parte Florida Mental Health Institute, University of Florida; 2005.
10. Godley MD, Godley SH, Dennis ML, Funk RR, Passetti LL, Petry NM. A randomized trial of Assertive Continuing Care and Contingency Management for adolescents with substance use disorders. *Journal of Consulting and Clinical Psychology*. 2014; 82(1):40–51. [PubMed: 24294838]
11. Godley MD, Godley SH, Dennis ML, Funk RR, Passetti LL. The effect of assertive continuing care on continuing care linkage, adherence and abstinence following residential treatment for adolescents with substance use disorders. *Addiction*. 2007; 102(1):81–93. [PubMed: 17207126]
12. Godley, SH., Meyers, RJ., Smith, JE., Godley, MD., Titus, JC., Karvinen, T., Kelberg, P. The adolescent community reinforcement approach (ACRA) for adolescent cannabis users, Cannabis Youth Treatment (CYT) Manual Series, volume 4 (DHHS Publications No. [SMA] 01-3489). Rockville, MD: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration; 2001.
13. Green AE, Trott E, Willging CE, Finna NK, Ehrhardt MG, Aarons GA. The role of collaborations in sustaining an evidence-based intervention to reduce child neglect. *Child Abuse & Neglect*. 2015; 53:4–16. DOI: 10.1016/j.chiabu.2015.11.013 [PubMed: 26712422]
14. Henderson CE, Wevodau AL, Henderson SE, Colbourn SL, Gharagozloo L, North LW, Lotts VA. An independent replication of the adolescent community reinforcement approach with justice-involved youth. *American Journal on Addictions*. 2016; XX:1–8. DOI: 10.1111/ajad.12366
15. Hodge LM, Turner KM, Sanders MR, Filus A. Sustained implementation support scale: validation of a measure of program characteristics and workplace functioning for sustained program implementation. *Journal of Behavioral Health Services & Research*. 2016; doi: 10.1007/s11414-016-9505-z
16. Hupp, C., Mertig, K., Krall, K., Godley, MD., Godley, SH. Adolescent community reinforcement approach (A-CRA) and Assertive Continuing CARE (ACC) supervisor rating manual. Chestnut Health Systems; Normal, IL: 2009.
17. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*. 2015; 10(1):155. [PubMed: 26537706]
18. Martino S. Strategies for training counselors in evidence-based treatments. *Addiction Science & Clinical Practice*. 2010; 5(2):30–40. [PubMed: 22002451]
19. McGovern MP, Matzkin AL, Giard J. Assessing the dual diagnosis capability of addiction treatment services: The Dual Diagnosis Capability in Addiction Treatment (DDCAT) Index. *Journal of Dual Diagnosis*. 2007; 3(2):111–123.
20. Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*. 2004; 72(6):1050–1062. [PubMed: 15612851]
21. Proctor E, Luke D, Calhoun A, McMillen C, Brownson R, McCrary S, Padek M. Sustainability of evidence-based healthcare: research agenda, methodological advances, and infrastructure support. *Implementation Science*. 2015; 10(1):88. [PubMed: 26062907]
22. Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: MESA; 1960.
23. Reckase, MD. Multidimensional item response theory. New York: Springer; 2009.
24. Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
25. SAS Institute Inc. SAS/STAT® 14.1 User's Guide. Cary, NC: SAS Institute Inc.; 2015.
26. Scheirer MA, Dearing JW. An agenda for research on the sustainability of public health programs. *American Journal of Public Health*. 2011; 101(11):2059–2067. [PubMed: 21940916]

27. Scheirer MA, Hartling G, Hagerman D. Defining sustainability outcomes of health programs: Illustrations from an on-line survey. *Evaluation and Program Planning*. 2008; 31(4):335–346. [PubMed: 18835642]
28. Schoenwald SK, Sheidow AJ, Chapman JE. Clinical supervision in treatment transport: effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*. 2009; 77(3):410–421. [PubMed: 19485583]
29. Simpson DD, Flynn PM. Moving innovations into treatment: A stage-based approach to program change. *Journal of Substance Abuse Treatment*. 2007; 33(2):111–120. [PubMed: 17434704]
30. Slesnick N, Prestopnik JL, Meyers RJ, Glassman M. Treatment outcome for street-living, homeless youth. *Addictive Behavior*. 2007; 32(6):1237–1251.
31. Smith, JE., Lundy, SL., Gianini, L. Community Reinforcement Approach (CRA) and Adolescents Community Reinforcement Approach (A-CRA) Therapist Coding Manual. Chestnut Health Systems; Normal, IL: 2007.
32. Wiltsey Stirman WS, Kimberly J, Cook N, Calloway A, Castro F, Charns M. The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. *Implementation Science*. 2012; 7(17)doi: 10.1186/1748-5908-7-17

Highlights

- This study operationalizes the extent of behavioral treatment sustainment
- Different aspects of organizational support are defined
- Modern psychometric assessment tools are rarely used in health care to assess sustainment

Figure 1.
Treatment staffing and delivery (Component #1) scores by dichotomous sustainment status

Figure 2.
Comparing Supervision (Component #2) scores by dichotomous sustainment status.

Table 1

Descriptive statistics of the sustainment elements from participating sites

Sustainment Elements (scale values or scale range)	Data distribution	
	<i>Mean (SD) or mean % (range)</i>	<i>% meeting criteria</i>
1) Clinical knowledge of A-CRA (10 items; factor score)	0 (0.57) *	
2) Presence of certified A-CRA clinician(s) (0, 1)		81.97%
3) Presence of a certified A-CRA supervisor(s) (0, 1)		63.93%
4) A-CRA training and certification process knowledge (0 – 100%)	53.76% (0–100%)	
5) Percentage of youth receiving A-CRA (0 – 100%)	49.15% (0–100%)	
6) A-CRA dosage (0, 1)		66.67%
7) Clinical supervision frequency (0, 1)		50.94%
8) Clinical supervision content (6 items; factor score)	0 (0.33) *	
9) Supervision with treatment session review (0, 0.5, 1)	0.63 (0.45)	
10) Onsite A-CRA clinical training quality (0 – 1)	0.12 (0.26)	

* Note: IRT factor scores follow an underlying distribution of factor mean of 0 and standard deviation of 1. The sample mean is 0, meaning that the sample is at the average level of the assumed distribution of the underlying construct.

Table 2

Principal Components Analysis results

Elements:	Factor loading estimate		Standardized Scoring Coefficients	
	Component 1: Staffing & Treatment Delivery	Component 2: Supervision	Component 1: Staffing & Treatment Delivery	Component 2: Supervision
1) Clinical knowledge of A-CRA (10 items; factor score)	0.35	-0.51	0.12	-0.27
2) Presence of certified A-CRA clinician(s) (0, 1)	0.80	0.03	0.27	0.02
3) Presence of a certified A-CRA supervisor(s) (0,1)	0.75	0.27	0.25	0.14
4) A-CRA training and certification process knowledge (0 – 100%)	0.79	-0.13	0.27	-0.07
5) Percentage of youth receiving A-CRA (0 – 100%)	0.78	0.08	0.26	0.04
6) A-CRA dosage (0, 1)	0.49	-0.39	0.17	-0.21
7) Clinical supervision frequency (0, 1)	-0.22	0.58	-0.08	0.30
8) Clinical supervision content (6 items; factor score)	0.13	0.66	0.05	0.35
9) Supervision with treatment session review (0, 0.5, 1)	0.22	0.71	0.08	0.37
10) Onsite A-CRA clinical training quality (0 – 1)	0.20	0.33	0.07	0.18

Note: there are three levels of shading indicating low (< 0.4), moderate (0.4 to 0.55) and high (> 0.55) levels of influence of each element on the corresponding principal component.