Sustaining an evidence-based program over time: Moderators of sustainability and the role of the Getting To Outcomes implementation support intervention

Joie Acosta, Ph.D.[*],
RAND Corporation

Matthew Chinman, Ph.D.,
RAND Corporation

Patricia Ebener, BS,
RAND Corporation

Patrick S. Malone, Ph.D.,
Malone Quantitative

Jill Cannon, Ph.D.,
RAND Corporation

Elizabeth J. D'Amico, Ph.D.
RAND Corporation

Abstract

Problematic rates of alcohol, e-cigarette and other drug use among US adolescents highlight the need for effective implementation of evidence-based prevention programs (EBPs), yet schools and community organizations have great difficulty implementing and sustaining EBPs. Although a growing number of studies show that implementation support interventions can improve EBP implementation, the literature on how to improve sustainability through implementation support is limited. This randomized controlled trial advances the literature by testing the effects of one such implementation intervention—Getting To Outcomes (GTO)—on sustainability of CHOICE, an after-school EBP for preventing substance use among middle-school students. CHOICE implementation was tracked for two-years after GTO support ended across 29 Boys and Girls Club

[*]Corresponding author, jacosta@rand.org.

sites in the greater Los Angeles area. Predictors of sustainability were identified for a set of key tasks targeted by the GTO approach (e.g., goal setting, evaluation, collectively called 'GTO performance') and for CHOICE fidelity using a series of path models. One year after GTO support ended we found no differences between GTO and control sites on CHOICE fidelity. GTO performance was also similar between groups, however GTO sites were superior in conducting evaluation. Better GTO performance predicted better CHOICE fidelity. Two years after GTO support ended, GTO sites were significantly more likely to sustain CHOICE implementation when compared with control sites. This study suggests that using an implementation support intervention like GTO can help low-resource settings continue to sustain their EBP implementation to help them get the most out of their investment.

Keywords

implementation support; evidence-based program; sustainability; adolescents; intervention

Introduction

Effective implementation of evidence-based prevention programs is needed given the problematic rates of alcohol, marijuana, and other drug use among US adolescents. In 2018, over half of high school seniors reported drinking alcohol, and one third reported being drunk in the past year. One third of seniors acknowledged drinking in the past month, and over 20% reported using marijuana in the past month. Electronic cigarette use among youth is now greater than cigarette smoking, with two out of every five 12th graders reporting past-year vaping (Johnston et al., 2019). The estimated costs of alcohol misuse, illicit drug use, and substance use disorders are more than $400 billion (Drug Enforcement Agency, 2017; Sacks et al., 2015) per year.

Support interventions improve implementation of evidence-based programs

Communities often face difficulty implementing evidence-based programs (EBPs) with the quality needed to achieve outcomes. A recent review of over two decades of studies on drug prevention EBP implementation shows a consistently low rate of EBP adoption and poor fidelity in schools and community organizations (Chinman et al. 2019), despite the availability of scores of EBPs disseminated by program developers and registries (e.g., Blueprints for Healthy Youth Development). Poor implementation often results from limited resources and a lack of the knowledge, attitudes, and skills (defined as *capacity*) that practitioners need to implement "off-the-shelf" EBPs (Wandersman and Florin 2003). Getting To Outcomes (GTO) is an evidence-based intervention that builds capacity for implementing EBPs by strengthening the knowledge, attitudes, and skills needed to choose, plan, implement, evaluate, and sustain EBPs. GTO lays out 10 key steps (Table 1) for obtaining positive results that reflect implementation best practices: Steps 1–6 address planning EBPs, Steps 7–8 involve process and outcome evaluation, and Steps 9–10 focus on the use of data to improve and sustain programs (Acosta et al. 2013; Chinman et al. 2009; Chinman et al. 2018a; Chinman et al. 2018b). Three types of implementation supports are provided to help organizations progress through the 10 steps: a manual (Wiseman et al. 2007) that is specifically tailored to alcohol and drug prevention programming, face-to-face

staff training, and onsite technical assistance. These implementation supports are comprised of several implementation strategies (Powell et al., 2015)—the GTO technical assistance is similar to "facilitation" in the implementation science literature (Kirchner et al., 2014). Like facilitation, GTO technical assistance emphasizes change in work practices through encouragement and action promotion via regular, ongoing meetings (Rycroft-Malone et al., 2004). GTO training and technical assistance providers guide practitioners to use GTO-based tools to adapt and tailor the implementation of an EBP (e.g., GTO Step 4) using multiple evaluative (e.g., GTO Steps 7–8) and iterative (e.g., GTO Step 9) strategies. Together these implementation supports help build a practitioner's capacity to use the implementation best practices specified by the GTO 10-steps.

Less is known about how to ensure effective evidence-based programs are sustained

Despite the plethora of information on predictors of sustainability, the literature on how to improve sustainability through implementation support, such as PROmoting School-communityuniversity Partnership to Enhance Resilience or Communities That Care, is limited. Consistent factors have not been tracked across the literature, nor has the literature reached a consensus on the most relevant sustainability outcomes to measure. Prior research has defined sustainability in a variety of ways: sustaining any part of an EBP, sustaining with fidelity, and sustaining with fidelity plus using implementation best practices (Hailemariam et al., 2019; Johnson et al., 2017). However, there is a growing consensus towards defining sustainability in a multi-dimensional manner: "(1) after a defined period of time, (2) the program, clinical intervention, and/or implementation strategies continue to be delivered and/or (3) individual behavior change (i.e., clinician, patient) is maintained; (4) the program and individual behavior change may evolve or adapt while (5) continuing to produce benefits for individuals/ systems" (Moore, Mascarenhas, Bain, & Straus, 2017). A review of 125 studies found the most common influences of sustainability were workforce stability, strong implementation, and adequate funding (Stirman et al. 2012); however, there was no single pathway to sustainability (Welsh et al., 2016). Research also found that fidelity and outcome monitoring, ongoing supervision (Peterson et al. 2014), knowledge of logic models, communication with trainers or program developers, sustainability planning, and alignment with the goals of the implementing agency all significantly predicted sustainability of an EBP (Cooper et al. 2015).

Research Questions and Study Hypotheses

This study builds on a prior work conducted as part of a two-year randomized controlled trial of GTO (Chinman et al. 2018). The prior work provides an opportunity to explore specific predictors of a site's ability to sustain all or some elements of CHOICE—an after-school EBP for preventing substance use among middle-school students—after GTO support ends. We found that after the two years of support, GTO sites had improved CHOICE adherence and quality of delivery whereas sites not assigned to GTO had not improved. Also, GTO sites had higher ratings of quality on how they performed key programming tasks for each GTO step (i.e., GTO performance; Chinman et al. 2018a). A study timeline is available in Figure S1 of the supplemental material. In this study, we built on this prior work and the sustainability literature by examining two research questions focused on what

happens in the first and second years after GTO support ended: **1. Did more GTO sites sustain delivery of an evidence-based program (i.e., CHOICE) than control sites? Hypothesis 1.1.** After a year without GTO support, GTO sites will sustain their improvements—specifically, GTO sites will have better GTO performance and better CHOICE fidelity when compared with control sites; and **Hypothesis 1.2.** After a year without GTO support, more GTO sites will continue to implement CHOICE compared to control sites. **2. What predicted the sustainability of the implementation gains made by GTO sites (i.e., improvements in performance, CHOICE delivery)? Hypothesis 2.1.** After a year without GTO support, GTO dosage (i.e., hours of GTO assistance utilized), CHOICE training and experience, organizational support for evidence-based practices, and implementation barriers and facilitators will predict ratings of GTO performance and CHOICE fidelity. **Hypothesis 2.2.** After two years without GTO support, GTO performance will predict concurrent CHOICE fidelity and sites' continuation of CHOICE.

For this study we define sustainability as continuing to implement CHOICE, delivering CHOICE with fidelity, and continuing to implement the implementation best practices recommended by GTO (measured as GTO performance). The study advances the existing literature. First, the study was conducted prospectively—i.e., measures were used with the express intent of measuring sustainability rather than using data initially designed for another purpose. This type of prospective, multi-level, mixed method design is ideal for studying sustainability (Shelton, Cooper & Stirman, 2018). Second, the measures address key sustainability outcomes (e.g., continuation of CHOICE, ongoing fidelity) important for implementation. Third, we examined several predictors of sustainability, including both implementation factors (e.g., barriers and facilitators) and implementer characteristics (e.g., staff turnover, training).

Methods

Study Design

We compared two groups of Boys & Girls Clubs sites implementing the CHOICE intervention twice over a two-year period—with (intervention, n=14 sites) and without (control, n=15 sites) GTO assistance. Intervention and control groups were assessed on their GTO performance, CHOICE fidelity (e.g., adherence, quality of delivery) and the alcohol and drug outcomes of participating middle schoolers (Chinman et al. 2018b). One of the GTO intervention sites closed between Year 2 and 3 (n=13 GTO sites for Year 3 analyses).

Participants

Twenty-three sites were in Los Angeles, California, and six were in neighboring Orange County (Boys and Girls Clubs can operate multiple sites). More information about the participants and the randomization process is available in Chinman et al. (2018a). The site-level sample size was justified at 80% power by taking into account the estimated correlation between baseline and follow-up assessments of the site-level measures (0.5 to 0.6) and the moderate to large effect sizes expected based on previous GTO studies (Chinman et al. 2016; Chinman et al. 2018a).

CHOICE: An evidence-based alcohol and drug prevention program

CHOICE involves five, 30 minute sessions based on social learning, decision-making and self-efficacy theories and has been associated with reductions in alcohol and marijuana use (D'Amico et al. 2012). Two half-time, master's-level technical assistance providers delivered standard CHOICE manuals and training to all sites. At GTO sites, technical assistance providers also delivered GTO manuals, face-to-face training, and onsite technical assistance with phone and email followup to support implementation of CHOICE delivery. Table 1 shows how site staff used GTO to implement CHOICE. Full details are described elsewhere (Chinman et al. 2018b).

Measures and Data Collection

This study was approved by RAND's Institutional Review Board. Data collectors and technical assistance staff watched for harms during the study. None were reported. All sites received $2,000 to defray the cost of participating in the study. Data described below were collected at different times throughout the study (Figure S1). Boys and Girls Club site staff demographics and background were collected at baseline; GTO dosage was collected in Years 1 and 2 (while GTO was active); GTO performance and fidelity were collected in Years 1–3, with Year 3 being one year after GTO was ended; implementation characteristics were collected in between Years 2 and 3; and the continuation of CHOICE was assessed in Years 3 and 4 (one and two years after GTO was ended, respectively).

GTO dosage—For the two years of GTO support, each technical assistance provider logged their time, by GTO step, spent delivering training and technical assistance to site staff (hours of technical assistance per GTO site, Year 1: $M = 11.17$, $SD = 3.4$; Year 2: $M = 14.7$, $SD = 3.9$).

GTO performance—As in past GTO studies (Chinman et al. 2008a; Chinman et al. 2009), we used the structured Performance Interview annually with staff members responsible for running CHOICE. Ratings are made at the site level because programs operate as a unit. The interview consisted of 12 items that assessed how well sites performed key tasks in eight domains (i.e., aligned with eight GTO steps) throughout CHOICE implementation (e.g., developing goals and desired outcomes, ensuring program fit). Responses to each item were rated on a five-point scale from "highly faithful" to ideal practice (=5) to "highly divergent" from ideal practice (=1), guided by by specific criteria. The ratings for each domain were combined into a measure that yielded a score for each domain and a total overall performance score. More detail on data collection and reliability of this measure is available in Chinman et al. (2018b). In Year 4, the instrument had relatively few questions and a simpler set of response options (e.g., Run CHOICE in the last 6 months? [response options: yes/no], Collect any fidelity or outcome evaluation data? [response options: neither, both, fidelity only, outcomes only]). Thus, we did not compute alpha, but 67% of double-coded responses were exact matches.

CHOICE fidelity—All sites were rated on *adherence* to the CHOICE protocol and *quality* of delivery by a pool of eight data collectors (blind to condition). To calculate reliability the data collectors rated 16 videotaped sessions. Krippendorff's α was calculated comparing

observers' ratings for each video to the "master ratings" by the CHOICE trainer. Fidelity data were collected each year (Years 1–3), but this study only used data from Years 2 and 3 (see Figure S1 for a complete study timeline). (1) *Adherence*: Data collectors observed and rated two randomly selected CHOICE sessions per site on how closely site staff implemented activities as designed (not at all, partially, or fully) using a CHOICE fidelity tool (D'Amico et al. 2012). In Year 1, we rated 489 activities (36%) distributed across all 29 sites (n=235 for the control group, n=254 for the intervention group). In Year 2, we rated 515 activities (38%), distributed across all 29 sites (n=255 for the control group, 260 for the intervention group). Ordinal α comparing ratings from each of the eight coders to the master key ranged from 0.50 to 0.91, *Median* = 0.70, acceptable to good by common standards (Krippendorff 2004). (2) *Quality of delivery:* The Motivational Interviewing Treatment Integrity scale (Moyers et al. 2010) was used to assess five specific behaviors that are counted during the session and five "global" ratings in which the entire session is scored on a scale from 1 (low) to 5 (high). This scale has shown acceptable psychometric characteristics across multiple research settings (Campbell et al. 2009; Turrisi et al. 2009), and its scores have correlated with outcomes as expected, suggesting its validity (Pollak et al. 2014). However, because four of the five global ratings had low inter-rater reliability (α < 0.55) in this study, we retained only "evocation" (estimated Krippendorff's α = 0.65) and otherwise relied on counts of specific behaviors to operationalize delivery quality. The behaviors counted during the sessions are the number of open- and closed-ended questions, statements that are Motivation Interviewing-adherent (e.g., "If it's ok, I'd like to hear what you think.") or non-adherent (e.g., "You need to stop using drugs"), and reflections that are simple (e.g., "Some of you are ready to make changes") or complex (e.g., "Some of you are hoping that by making changes, things will improve in your lives"). From these counts, we derived: percent complex reflections, open questions, Motivational Interviewing-adherent statements, and reflection to question ratio. α for each data collector was high (from 0.88 to 0.93, *Median* = 0.90; Cicchetti 1994).

Boys and Girls Club site staff demographics and background—At baseline (after randomization), we conducted a web-based survey of site staff involved in CHOICE to assess differences in demographic variables and organizational support for evidence-based practices. All staff responded (control = 29; intervention = 34). Staff in the control and intervention groups had similar demographic makeup (no significant differences based on bivariate models accounting for clustering within Boys and Girls club and the county). Details on staff demographics can be found in Chinman et al. (2018b). The web survey included the Organizational Support for Evidence-Based Practices Scale (Aarons et al. 2009), a nine-item scale to assess processes and structures supporting the use of evidence-based practices within an organization (e.g., agency sponsored trainings or in-services). Its McDonald's ω[1] coefficient (with 95% confidence intervals [CIs]) = 0.85, CI [0.74, 0.91]. The ω value found here is considered good (Krippendorff 2004). To evaluate baseline group differences on the scale, we fit a linear mixed effects regression model with fixed treatment effect (intervention vs. control) and random intercepts for club, modeling county (Los

---

[1]Coefficient ω is a measure of internal consistency on the same scale as coefficient alpha but is less biased, has fewer problems than alpha, and has CIs to more accurately evaluate reliability (Dunn et al. 2014).

Angeles vs. Orange) as a higher-order level of stratification. The two groups did not differ significantly on the scale at baseline, $ps > 0.2$, with or without staff-level demographic covariates.

Implementation characteristics—We used a semi-structured interview guide based on the Consolidated Framework for Implementation Research (CFIR) to assess each site's implementation barriers and facilitators (referred to in this study as implementation characteristics, Table 2). All interviews were conducted by phone by one researcher, lasted approximately 45–60 minutes, and were audio recorded and transcribed verbatim. For each site, at least two staff members were invited to discuss their participation in CHOICE implementation. We interviewed 51 staff across the 29 sites (28 staff from intervention sites and 23 from control sites). Interviews occurred in June 2015 through July 2016 after all sites had run the CHOICE program for the second time, which corresponded to the end of GTO support. Final ratings were based on the consensus score from two raters on: 1) valence (either facilitating or hindering implementation) and 2) strength (the degree to which implementation was facilitated or hindered; Damschroder and Lowery 2013). The rating scale ranged from +2 (most facilitating) to −2 (most hindering). A zero rating reflected a neutral (e.g., lacking sufficient information) or mixed influence. More details on how CFIR was used in this project can be found in Cannon et al. (2019).

Statistical Analyses

Predictors of sustainability of implementation gains from GTO support—We explored predictors of sustainability in GTO sites using a series of path models testing different hypotheses 1.1 and 1.2 (predictors listed in Table 3). We began by establishing a graphical causal model (GCM; Elwert 2013) of our theorized relations among study variables as well as their unmeasured causes. Elwert provides a didactic presentation of the construction and use of GCMs, but in brief, a GCM is similar in appearance to a theoretical path model, but is not used as an analytic tool itself. Instead, it can be analyzed with tools such as *DAGitty* (Textor, van der Zander, Gilthorpe, Liskiewiecz, & Ellison, 2016) to determine what paths among variables meet the causal-inference criteria of the Structural Causal Model (SCM; Pearl, 2011), given certain sets of confounding variables. We used these graphical methods to identify the covariate sets (where they exist) required for causal interpretation between the given cause of interest and given effect (sustained GTO performance or sustained CHOICE fidelity). Thus we can assert that observed relations, when the covariates are controlled are causal given a valid GCM. It should be noted that the GCM is not necessarily directly estimated and is therefore not constrained to any specific estimation technique or functional form of the relations (Pearl 2011). We used the GCM to identify specific testable submodels and estimated them as linear path models using maximum likelihood in M*plus*. The current use of linearity and the ML estimator should be noted as a caveat to the strong causal interpretation. Finally, we applied the Benjamini-Hochberg (Benjamini & Hochberg 1995) procedure to the multiple tests for each outcome to constrain the false discovery rate to 0.05.

GTO's influence on CHOICE sustainability—For Hypotheses 2.1 and 2.2, we analyzed intent-to-treat effects of experimental assignment on (1) GTO performance for

GTO Steps 7–10 and CHOICE fidelity (i.e., one year after the GTO intervention ended) and on (2) CHOICE implementation (i.e., two years after the GTO intervention ended), using linear and logistic regression models, with Fisher's exact tests for 2×2 tables.

Results

Descriptive statistics are shown in Table 3 for the study outcomes and the correlations among variables used in the causal analyses.

Hypothesis 1.1

After a year without GTO support, GTO dosage, CHOICE training and experience, organizational support for evidence-based practices, and implementation barriers and facilitators will predict ratings of GTO performance and CHOICE fidelity. Figure 1 provides the graphical casual model underlying our tests of the causes of sustained overall GTO performance and CHOICE fidelity. Using DAGitty v.2.3 (Textor, van der Zander, Gilthorpe, Liskiewiecz, & Ellison, 2015), we identified covariate sets for each cause/effect combination, where one existed. We tested five hypothesized causes of overall GTO performance in Year 3 (i.e., one year after GTO support ends): Year 1 and 2 GTO dosage (technical assistance hours utilized by each site); whether staff had experience teaching CHOICE by the end of the first and second year of study (yes/no); whether staff received the CHOICE training in the first and second year of study (yes/no); organizational support for evidence-based practice and GTO knowledge prior to intervention taken together (baseline; means score); and an aggregate of CFIR ratings in the second year of study (average rating from −2 to +2). We were able to establish covariates sets given our graphical casual model for all hypothesized causes except CFIR. The implied models are shown with standardized path coefficients in Figures S2 through S6 in supplemental material. Table 4 provides unstandardized regression coefficients with confidence intervals for the five key predictors. After false discovery rate correction, the data supported lower GTO dosage in Year 2 as a cause of higher subsequent overall Year 3 performance, $b = -0.030$, 95% CI [−0.054, −0.006], effect size $\beta = -0.61$. The results also show a positive relation between CFIR ratings (Year 2) and Year 3 GTO performance, $b = 0.640$, 95% CI [0.411, 0.879], $\beta = 0.78$, suggesting that having fewer implementation barriers and more facilitators predicted better GTO performance.

Hypothesis 1.2

After a year without GTO support, GTO performance will predict concurrent CHOICE fidelity and sites' continuation of CHOICE. We tested overall GTO performance, after two years without GTO support (i.e., in Year 3), as a cause of concurrent CHOICE fidelity in Year 3. CHOICE fidelity was measured by a mean of the five Motivational Interviewing Treatment Integrity global rating scores, the evocation rating by observers, and the adherence ratings. We also tested Year 3 GTO performance as a cause of sites' continuation of CHOICE in Year 3 (i.e., one year after GTO support ended). Because adherence observations were nested within sites, we estimated a two-level model of Year 3 GTO performance (level-2) effects on session observations (level-1) as well as on the Motivational Interviewing Treatment Integrity global ratings and evocation (both level-2).

We established GTO performance as a cause from the GCM. The implied models with standardized path coefficients are in Figure S6 in the supplemental material. Table 5 presents standardized and unstandardized results. After false discovery rate correction, overall GTO performance in Year 3 was supported as a cause of the Year 3 Motivational Interviewing Treatment Integrity mean global rating score, $b = 0.16$, 95% CI [0.07, 0.26], $\beta = 0.68$.

Hypothesis 2.1

One year after GTO support ends, GTO performance and CHOICE fidelity is greater among GTO sites compared with control sites. We conducted intent-to-treat analyses for eleven measures in separate models. Year 3 GTO performance was measured by progress in four of the GTO steps: Steps 7 (process evaluation), 8 (outcome evaluation), 9 (quality improvement), and 10 (sustainability). The earlier GTO steps were not tested because sites' focus was evaluation, improvement, and sustainability at that time. Each step was coded as a binary variable: "1" indicated that improvement had occurred since the end of support or that satisfactory performance had already been achieved by the end of support, or "0" = otherwise. These were tested using Fisher's exact test for 2×2 tables, given the modest sample size (Table 6). After false discovery rate correction across the 4 Fisher tests, only Step 8 showed significantly better performance for GTO (9 of 9 sites, 100%) compared to control (1 of 13 sites, 8%), $\phi = 0.91$. This means that GTO sites were more likely to be conducting outcome evaluation than control sites.

Year 3 CHOICE fidelity was measured using the observer rating of evocation in CHOICE sessions and the computation of four derived variables: percent complex reflections, percent open questions, reflection question ratio, percent motivational interviewing-adherent. We also measured fidelity with number of CHOICE sessions held (dichotomized to five or less than five) and adherence ratings to the protocol. We modeled evocation and the four Motivational Interviewing Treatment Integrity scale-derived variables as intervals. We estimated number of sessions using a logistic model. Finally, we estimated the session-level ratings in a two-level cumulative logistic model covarying specific session and allowing random intercepts by site (Table 6). We did not find statistically significant differences on any of the fidelity measures, ps > 0.05.

Hypothesis 2.2

Two years after GTO support ends, more GTO sites continue to implement CHOICE than control sites. We conducted an intent-to-treat analysis modeling whether or not a site continued to use CHOICE in Year 4 (i.e., two years after GTO support ends). Six of 13 (46%) GTO sites continued CHOICE; none of the 12 reporting control sites did. This was a significant difference by Fisher's exact test, $p = 0.015$.

Discussion

This study found that GTO intervention sites attempting a substance use prevention EBP had greater fidelity and GTO performance by the end of the two-year GTO intervention (Chinman et al. 2018b). One year after GTO support ended (Year 3), intervention sites sustained their superior performance on outcome evaluation (Step 8), but not on other GTO

steps (7-process evaluation, 9continuous quality improvement, 10-sustainability). Also, there were no significant differences between intervention and control sites on CHOICE fidelity in Year 3, suggesting GTO sites were not able to sustain their initial improvements in fidelity. However, two years after GTO ended (i.e., Year 4), intervention sites were more likely to have sustained the EBP (i.e., CHOICE). These findings suggest that GTO helps sites sustain an EBP and outcome evaluation of the EBP, although implementation quality may decline without some ongoing GTO support.

Significant predictors of overall GTO performance in Year 3 included lower levels of GTO dosage (technical assistance hours) and more positive implementation characteristics (i.e., fewer barriers and more facilitators as measured by CFIR). It is unclear why lower GTO dosage predicted higher GTO performance in Year 3, however it may be that some sites who used less technical assistance hours did so because they did not need the additional support to attain a high level of performance, because they had received sufficient GTO support in the prior year. Year 3 overall GTO performance was a significant predictor of CHOICE fidelity in Year 3 across all sites. This association is important to document because it underlies the model of how implementation support works (reflected in how well sites perform GTO-targeted activities) and replicates the same finding in an earlier GTO study (Chinman et al. 2016). This builds the evidence base for implementation support as an effective intervention to achieve fidelity among EBPs and is consistent with prior research on coalitions that found that implementation skills and planning processes were associated with sustainability (Shelton, Cooper & Stirman, 2018). It is often difficult for low-resource organizations to continue implementing effective programs with fidelity given limited capacity and changes in staffing and funding (Stirman et al. 2012; Peterson et al. 2014). This study demonstrates that GTO implementation support helped about half of the low-resource community-based settings maintain implementation of an prevention EBP two years after outside GTO support ended. Findings are consistent with prior research that found that implementation supports improve EBPs sustainability (Spoth et al. 2011; Gloppen, Arthur, Hawkins, & Shapiro 2012). Similar to the findings of Johnson et al. (2017), our study also found that implementation facilitators and barriers predicted sustainability (e.g., trialability). In contrast, fidelity gains made after the GTO intervention were not maintained one year later, suggesting some ongoing support may be needed to maintain higher levels of fidelity.

Study Limitations

Data collection was limited to two years after GTO support ended. Although the use of a single program across similar settings is a strong experimental design, it could limit generalizability of findings beyond low-resource community-based settings. Although we explicitly defined sustainability for the context of this study (as recommended by Hailemariam et al. 2019), the definition was limited to EBP fidelity and performance of key implementation best practices. Finally, participant outcomes were not tracked past when GTO support ended.

Implications and Future Directions

In order to broadly diffuse EBPs, programs and practices need to not only be implemented while there is support – but sustained after support ends. Future research needs to explore

models, like GTO, that can help to promote sustainability of high quality EBP implementation over time. Future GTO research should focus more in-depth on GTO Step 9 (continuous quality improvement), which can help sites determine whether there is a continued need for the EBP to be implemented over time, and when adaptations might be needed. Adaptation is a key—yet understudied component of sustainability (Moore et al., 2017; Shelton & Lee, 2019). This study found that implementation support strategies like GTO can aid implementation and sustainability. We found that *some* implementation support is needed to sustain fidelity over time. Future research is needed to identify the appropriate *level* of implementation support to maintain high fidelity. For example, given the investment in time (approximately two hours a month) and impact of GTO on fidelity is it cost effective to continue GTO implementation support over time? Understanding how to best structure implementation support in a cost effective manner will be necessary to scale up these supports, which can identify adaptations needed to improve EBP fit over time (Shelton, Cooper & Stirman, 2018). This study contributes to the evidence base by not only looking at continuing core components of the CHOICE intervention as an outcome, but also at the extent to which specific implementation practices predict sustainability. Also, using CFIR – a widespread implementation framework—we tested the extent to which specific implementation characteristics predict sustainability. Although CFIR ratings were not found to significantly predict sustainability, they were correlated with higher ratings of sites' performance of implementation best practices (i.e., GTO performance). Future research should continue testing whether CFIR is a useful conceptual framework for supporting sustainability (Shelton & Lee, 2019). In addition to examining the sustainability of implementation quality, future research should also examine the sustainability of outcomes at the individual level to determine how outcomes of an EBP and its overall public health impact can be sustained over time. Cost-effectiveness studies of implementation support approaches like GTO should also test the lasting effects on implementation after the support ends. Sustaining EBPs is a critical step in "scaling up" EBPs in a manner likely to have a measurable impact on today's critical social problems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aarons GA, Sommerfeld DH, & Walrath-Greene CM (2009). Evidence-based practice implementation: the impact of public versus private sector organization type on organizational support, provider attitudes, and adoption of evidence-based practice. Implement Sci, 4, 83, doi:10.1186/1748-5908-4-83. [PubMed: 20043824]

Acosta J, Chinman M, Ebener P, Malone PS, Paddock S, Phillips A, et al. (2013). An intervention to improve program implementation: findings from a two-year cluster randomized trial of Assets-Getting To Outcomes. Implementation science : IS, 8, 87–87, doi:10.1186/1748-5908-8-87. [PubMed: 23924279]

Bandura A (2004). Health promotion by social cognitive means. Health Educ Behav, 31(2), 143–164, doi:10.1177/1090198104263660. [PubMed: 15090118]

Campbell MK, Carr C, DeVellis B, Switzer B, Biddle A, Amamoo MA, et al. (2009). A Randomized Trial of Tailoring and Motivational Interviewing to Promote Fruit and Vegetable Consumption for Cancer Prevention and Control. [journal article]. Annals of Behavioral Medicine, 38(2), 71, doi:10.1007/s12160-009-9140-5. [PubMed: 20012809]

Cannon JS, Gilbert M, Ebener P, Malone PS, Reardon CM, Acosta J, & Chinman M (2019). Influence of an Implementation Support Intervention on Barriers and Facilitators to Delivery of a Substance Use Prevention Program. Prevention Science. 10.1007/s11121-019-01037-x.

Chinman M, Acosta J, Ebener P, Hunter S, Imm P, & Wandersman A (2019). Dissemination of Evidence-Based Prevention Interventions and Policies In Sloboda Z, Petras H, Robertson E, & Hingson R (Eds.), Prevention of Substance Use (pp. 367–383). Cham: Springer International Publishing.

Chinman M, Acosta J, Ebener P, Malone PS, & Slaughter ME (2016). Can implementation support help community-based settings better deliver evidence-based sexual health promotion programs? A randomized trial of Getting To Outcomes®. [journal article]. Implementation Science, 11(1), 78, doi:10.1186/s13012-016-0446-y. [PubMed: 27245158]

Chinman M, Acosta J, Ebener P, Malone PS, & Slaughter ME (2018a). A Cluster-Randomized Trial of Getting To Outcomes' Impact on Sexual Health Outcomes in Community-Based Settings. Prev Sci, 19(4), 437–448, doi:10.1007/s11121-017-0845-6. [PubMed: 28971273]

Chinman M, Ebener P, Malone PS, Cannon J, D'Amico EJ, & Acosta J (2018b). Testing implementation support for evidence-based programs in community settings: a replication cluster-randomized trial of Getting To Outcomes(R). Implement Sci, 13(1), 131, doi:10.1186/s13012-018-0825-7. [PubMed: 30348227]

Chinman M, Hunter SB, Ebener P, Paddock SM, Stillman L, Imm P, et al. (2008a). The Getting To Outcomes Demonstration and Evaluation: An Illustration of the Prevention Support System. [journal article]. American Journal of Community Psychology, 41(3), 206224, doi:10.1007/s10464-008-9163-2.

Chinman M, Hunter SB, Ebener P, Paddock SM, Stillman L, Imm P, et al. (2008b). The getting to outcomes demonstration and evaluation: an illustration of the prevention support system. American Journal of Community Psychology, 41(3–4), 206–224, doi:10.1007/s10464-008-9163-2. [PubMed: 18278551]

Chinman M, Tremain B, Imm P, & Wandersman A (2009). Strengthening prevention performance using technology: a formative evaluation of interactive Getting To Outcomes. The American journal of orthopsychiatry, 79(4), 469–481, doi:10.1037/a0016705. [PubMed: 20099938]

Cicchetti D (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology (Vol. 6).

D'Amico EJ, Tucker JS, Miles JNV, Zhou AJ, Shih RA, & Green HD Jr. (2012). Preventing alcohol use with a voluntary after-school program for middle school students: results from a cluster randomized controlled trial of CHOICE. Prev Sci, 13(4), 415–425, doi:10.1007/s11121-011-0269-7. [PubMed: 22311178]

Damschroder LJ, & Lowery JC (2013). Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci, 8, 51, doi:10.1186/1748-5908-8-51. [PubMed: 23663819]

Dedoose v.8.0.42, online database (2016). www.dedoose.com. Accessed 04/30/2016.

Drug Enforcement Administration. (2017). National drug threat assessment. Washington (DC): Department of Justice (US).

Dunn TJ, Baguley T, & Brunsden V (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol, 105(3), 399–412, doi:10.1111/bjop.12046. [PubMed: 24844115]

Dusenbury L, Brannigan R, Hansen WB, Walsh J, & Falco M (2005). Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. Health Educ Res, 20(3), 308–313, doi:10.1093/her/cyg134. [PubMed: 15522898]

Fetterman DM, & Wandersman A (2005). Empowerment evaluation principles in practice (Empowerment evaluation principles in practice.). New York, NY, US: Guilford Press.

Gloppen KM, Arthur MW, Hawkins JD, & Shapiro VB (2012). Sustainability of the Communities That Care prevention system by coalitions participating in the Community Youth Development Study. The Journal of adolescent health : official publication of the Society for Adolescent Medicine, 51(3), 259–264, doi:10.1016/j.jadohealth.2011.12.018. [PubMed: 22921136]

Hailemariam M, Bustos T, Montgomery B, Barajas R, Evans LB, & Drahota A (2019). Evidence-based intervention sustainability strategies: a systematic review. Implementation Science, 14(1), 57–68. [PubMed: 31171004]

Johnson K, Collins D, Shamblen S, Kenworthy T, & Wandersman A (2017). Long-Term Sustainability of Evidence-Based Prevention Interventions and Community Coalitions Survival: a Five and One-Half Year Follow-up Study. Prev Sci, 18(5), 610–621, doi:10.1007/s11121-017-0784-2. [PubMed: 28397156]

Johnston LD, Miech RA, O'Malley PM, Bachman JG, Schulenberg JE, & Patrick ME (2019). Monitoring the Future National Survey Results on Drug Use, 1975–2018: Overview, Key Findings on Adolescent Drug Use. Institute for Social Research.

Kirchner JE, Ritchie MJ, Pitcock JA, Parker LE, Curran GM, & Fortney JC (2014). Outcomes of a partnered facilitation strategy to implement primary care-mental health. J Gen Intern Med, 29(4), 904–12. doi:10.1007/s11606-014-3027-2. [PubMed: 25355087]

Krippendorff K (2004). Content Analysis: An Introduction to Its Methodology (2nd ed.). Sage. Organizational Research Methods, 13(2), 392–394.

Moore JE, Mascarenhas A, Bain J, & Straus SE (2017). Developing a comprehensive definition of sustainability. Implementation Science, 12(1), 110–117. [PubMed: 28865479]

Moyers T, Martin T, Manuel J, Miller W, & Ernst D (2010). Revised global scales: Motivational interviewing treatment integrity 3.1. 1 (MITI 3.1. 1). Unpublished manuscript, University of New Mexico, Albuquerque, NM.

Peterson AE, Bond GR, Drake RE, McHugo GJ, Jones AM, & Williams J (2014). Predicting the Long-Term Sustainability of Evidence-Based Practices in Mental Health Care: An 8-Year Longitudinal Analysis. The Journal of Behavioral Health Services & Research, 41(3), 337–346, doi:10.1007/s11414-013-9347-x. [PubMed: 23709327]

Pollak KI, Coffman CJ, Alexander SC, Østbye T, Lyna P, Tulsky JA, et al. (2014). Weight's up? Predictors of weight-related communication during primary care visits with overweight adolescents. Patient Education and Counseling, 96(3), 327–332, doi:10.1016/j.pec.2014.07.025. [PubMed: 25130793]

Powell BJ, Waltz TJ, Chinman MJ, Damschroder L, Smith JL, Matthieu MM, Proctor EK, & Kirchner JE (2015). A refined compilation of implementation strategies: Results from the expert recommendations for implementing change (ERIC) project. Implementation Science, 10, 21–34. [PubMed: 25889199]

Miller WR, & Rollnick S (2013). Motivational Interviewing: Helping People Change (3rd Edition). New York, NY, US: Guilford Press.

Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, & Titchen A (2004). An exploration of the factors that influence the implementation of evidence into practice. J Clinical Nursing, 13(8), 913–24. doi:10.1111/j.1365-2702.2004.01007.x.

Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. (2015). 2010 national and state costs of excessive alcohol consumption. American Journal of Preventive Medicine, 49(5), e73–e79. [PubMed: 26477807]

Shelton RC, Cooper BR, & Stirman SW (2018). The sustainability of evidence-based interventions and practices in public health and health care. Annual Review of Public Health, 39, 55–76.

Shelton RC, & Lee M (2019). Sustaining Evidence-Based Interventions and Policies: Recent Innovations and Future Directions in Implementation Science. Annual Review of Public Health, 39, 55–76.

Spoth R, Guyll M, Redmond C, Greenberg M, & Feinberg M (2011). Six-Year Sustainability of Evidence-Based Intervention Implementation Quality by Community-University Partnerships: The

PROSPER Study. [journal article]. American Journal of Community Psychology, 48(3), 412–425, doi:10.1007/s10464-011-9430-5. [PubMed: 21394561]

Stirman S, Kimberly J, Cook N, Calloway A, Castro F, & Charns M (2012). The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. Implement Sci, 7, 17, doi:10.1186/1748-5908-7-17. [PubMed: 22417162]

Textor J, van der Zander B, Gilthorpe MK, Liskiewicz M, & Ellison GTH (2015). Welcome to DAGitty [on-line]. Accessed November 6, 2018, from http://dagitty.net/

Textor J, van der Zander B, Gilthorpe MK, Liskiewicz M, & Ellison GTH (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty'. International Journal of Epidemiology, 45, 1887–1894. [PubMed: 28089956]

Turrisi R, Larimer ME, Mallett KA, Kilmer JR, Ray AE, Mastroleo NR, et al. (2009). A randomized clinical trial evaluating a combined alcohol intervention for high-risk college students. Journal of studies on alcohol and drugs, 70(4), 555–567. [PubMed: 19515296]

Wandersman A, & Florin P (2003). Community interventions and effective prevention. Am Psychol, 58(6–7), 441–448. [PubMed: 12971190]

Welsh JA, Chilenski SM, Johnson L, Greenberg MT, & Spoth RL (2016). Pathways to sustainability: 8-year follow-up from the PROSPER project. The Journal of Primary Prevention, 37(3), 263–286. [PubMed: 26892601]

Wiseman SH, Chinman M, Ebener PA, Hunter S, Imm PS, & Wandersman AH (2007). Getting To Outcomes: 10 steps for achieving results-based accountability. Santa Monica, CA: RAND.

Figure 1.
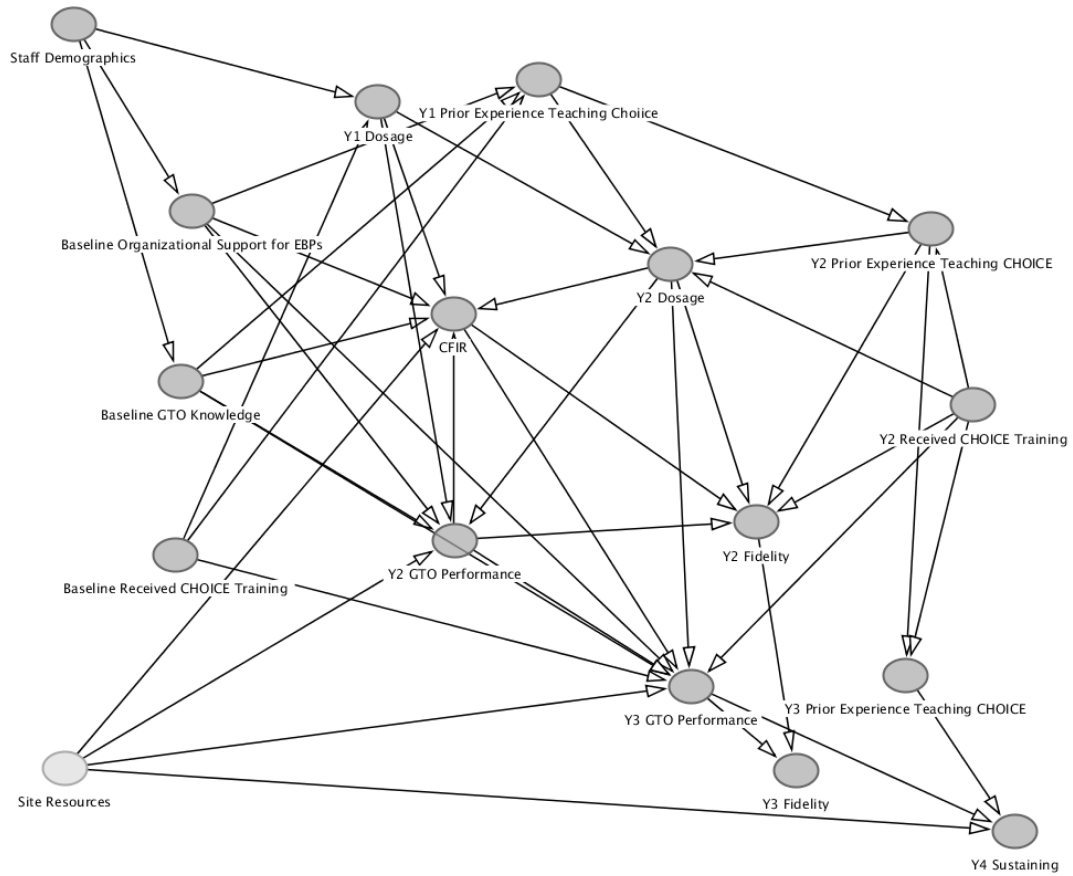Causal model (GCM) underlying our tests of causes of sustained GTO performance and
CHOICE fidelity in Year 3
Note: Y1 is Year 1, Y2 is Year 2, Y3 is Year 3 (one year after GTO ended) and Y4 is Year 4
(two years after GTO ended). Site resources was not a variable captured as part of this study,
but is included in this model because of its possible influence on GTO performance and
CHOICE fidelity.

Table 1.

How BGC staff in the intervention group performed various CHOICE implementation practices in each GTO step

GTO step	What the GTO manual provides for each step	Practices BGC club staff carried out within each GTO step
1. NEEDS: *What are the needs to address and the resources that can be used?*	Information about how to conduct a needs and resources assessment	Club staff reviewed data about the needs of their membership
2. GOALS AND OUTCOMES: *What are the goals and desired outcomes?*	Tools for creating measurable goals and desired outcomes	Each site developed its own broad goals and desired outcomes—statements that specify the amount and timing of change expected on specific measures of knowledge, attitudes, and behavior
3. BEST PRACTICES: *Which evidence-based programs can be useful in reaching the goals?*	Overview of the importance of using evidence-based programs and where to access information about them	Club leaders agreed to use CHOICE as the evidence-based program to implement
4. FIT: *What actions need to be taken so the selected program fits the community context?*	Tools to help program staff identify opportunities to reduce duplication and facilitate collaboration with other programs.	Each site reviewed CHOICE for how it would fit within the club and made adaptations to improve the fit
5. CAPACITY: *What capacity is needed for the program?*	Assessment tools to help program staff ensure there is sufficient organizational, human and fiscal capacity to conduct the program	Each site assessed its own capacity to carry out CHOICE and made plans to increase capacity when needed
6. PLAN: *What is the plan for this program?*	Information and tools to plan program activities in detail	Each site conducted concrete planning for CHOICE activities (e.g., who, what, where, when)
7. PROCESS EVALUATION: *How will the program implementation be assessed?*	Information and tools to help program staff plan and implement a process evaluation	Each site collected data on fidelity, attendance, and satisfaction to assess program delivery and reviewed that data immediately after implementation
8. OUTCOME EVALUATION: *How well did the program work?*	Information and tools to help program staff implement an outcome evaluation	Each site collected participant outcome data on actual behavior as well as on mediators such as attitudes and intentions
9. CONTINUOUS QUALITY IMPROVEMENT: *How will continuous quality improvement strategies be used to improve the program?*	Tools to prompt program staff to reassess GTO Steps 1–8 to stimulate program improvement plans	Each site reviewed decisions made and tools completed before implementation and data collected during and after implementation and made concrete changes for the next implementation
10. SUSTAINABILITY: *If the program is successful, how will it be sustained?*	Ideas to use when attempting to sustain an effective program	Each site considered ideas such as securing adequate funding, staffing, and buy-in to make it more likely that CHOICE would be sustained

Table 2.

CFIR constructs and examples of positive and negative indicators for the 17 constructs in this study

Construct and definition	Implementation facilitators	Implementation barriers
Innovation Characteristics Domain		
Relative advantage: perception better or worse than existing programs	Perceived by staff or leadership as being a better option for programming relevant to age group and subject matter	Perceived to be same as or worse than other drug and alcohol youth programming available
Adaptability: perception of ability to modify program to fit site's needs	Flexibility, inclusivity, and creative use of additional tools	Discomfort with and/or lack of ability to modify program
Complexity: how easy or hard the program is to deliver	Program perceived as easy, short, comfortable to deliver	Program perceived as difficult, overwhelming, and unfamiliar
Outer Setting Domain		
Needs and resources: extent to which participants' needs are known and prioritized	Good grasp of youths' needs and program adjusted to better suit them	Unfamiliarity with youths' needs or how programming can achieve better results
Inner Setting Domain		
Networks and communications: informal/ formal meetings at site	Frequent meetings, shared information to engage staff with programming	No meetings, discussions, or communications among leadership and staff around programming
Culture: consistency of staffing and site programming	Consistent staffing, little turnover, and programming info is passed on to new staff	Large turnover, burden on staff results in inconsistent programming
Implementation climate: general level of staff awareness/receptivity to program	Staff on board with and discuss new programs	No staff awareness and lack of acceptance for new programming
Compatibility: program fits within existing programming, mission, and time frame	Good fit with mission, staff qualities, experience, concerns of community, and timing of program	Not a good fit with staff or community and conflicts with other programming
Relative priority: shared perception of program's importance within site	Program importance is highlighted to community	Scheduling issues and/or lack of interest or lack of perceived need for program
Leadership engagement: leaders' commitment, involvement and accountability	Strong leadership, staff reported high commitment	No leadership support, leadership decentralized
Available resources: staff, space and time	Thorough, periodic review of staff and space resources	Inconsistent and/or unqualified staffing, no space
Access to knowledge and information: training on the mechanics of the program	Training and program information considered important for accurate delivery	Training did not occur and/or not all staff were trained, programming inconsistent
Implementation Process Domain		
Planning: pre-implementation strategizing and ongoing program-refining activities	Detailed and ongoing planning process with staff and leadership input	No planning or understanding of what planning is
Formally appointed internal implementation leaders (FIL): individuals responsible for delivering program	Leadership/staff teamwork, evidence of skill, experience, and engagement	Leadership and/or staff have no confidence in FIL, inconsistent staffing, poorly qualified FIL
Innovation participants: youth recruitment, engagement and retention strategies	Strategies are used to increase youth interest, participation, and attendance	Recruitment issues, lack of youth interest, and inconsistent attendance
Goal-setting*: program goals are communicated and acted upon by staff	Goal-setting was specific, reviewed, and changed over time to meet needs/ expectations	No goals established or discussed
Reflecting and evaluating: quantitative and qualitative feedback about progress and quality of program delivery	Reflecting and evaluation process used to control, shift focus, and make changes	No review of goals nor feedback given either during or after implementation

Note: Asterisk denotes a construct added to the original list of CFIR constructs.

Table 3.

Descriptive statistics for GTO sites

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.	Year 3, GTO Performance	1.0															
2.	Year 3, MITI	**0.66**															
3.	Year 3, Evocation	0.31	**0.54**														
4.	Year 3, CHOICE Adherence	0.46	0.41	**0.64**													
5.	Year 2, GTO Performance	**0.60**	0.28	0.48	**0.67**												
6.	Year 2, MITI	−0.05	0.19	0.18	0.14	−0.05											
7.	Year 2, Evocation	**0.44**	0.18	0.00	0.41	**0.55**	**−0.57**										
8.	Year 2, GTO Dosage	−0.17	−0.18	0.14	0.16	0.02	−0.32	0.54									
9.	Year 2, Prior Experience Teaching CHOICE	0.39	0.27	0.15	−0.05	0.17	0.03	−0.24	**−0.76**								
10.	Year 2, Received Choice Training	**−0.91**	**1.00**	1.00	1.00	**0.92**	−0.12	0.54	−0.04	0.83							
11.	Year 2, CFIR	**0.74**	**0.71**	**0.54**	**0.70**	**0.62**	−0.12	**0.53**	−0.02	0.17	1.00						
12.	Year 1, GTO Dosage	0.07	0.35	**0.41**	−0.13	−0.03	−0.09	0.13	**0.54**	−0.31	−0.23	0.12					
13.	Year 1, Prior Experience Teaching CHOICE	0.51	0.59	**0.58**	−0.12	−0.24	**0.99**	−1.00	−0.84	0.30	−0.48	0.12	0.34				
14.	Year1, Received Choice Training	0.28	0.54	**−0.58**	−0.66	−0.75	−0.88	0.54	0.66	**−0.64**	0.48	0.11	0.91	−0.48			
15.	Organizational Support for Evidence-based Practices	0.29	−0.22	0.08	−0.13	0.40	−0.33	0.06	−0.09	0.21	−0.39	0.13	0.15	0.83	−0.46		
16.	Baseline GTO Knowledge	0.19	0.22	−0.21	**−0.52**	−0.35	−0.11	0.00	0.02	0.18	−0.67	0.13	0.33	0.38	0.12	0.17	
N		13	12	12	12	13	12	12	13	13	13	13	11	13	13	13	13
Mean		3.4	0.6	3.0	0.8	3.4	0.7	4.0	21.4	1.3		0.4	18.7			0.7	4.7
SD		0.4	0.1	0.9	0.3	0.3	0.1	0.5	8.6	1.3		0.5	8.9			0.3	1.1

Note: Tabled values are correlation coefficients and other descriptive statistics with pairwise deletion for missing data. Bold type indicates a statistically significant correlation, $p < .05$.

Table 4.

Unstandardized and standardized causal coefficient estimates with confidence intervals for predictors of Year 3, GTO performance

Cause of Year 3, GTO performance	Causal coefficient (b)	Lower 95% CI	Upper 95% CI	z	Standardized coefficient (β)	Lower 95% CI	Upper 95% CI
Year 2, GTO dosage	−0.030[*]	−0.054	−0.006	−2.48	−0.61	−1.08	−0.13
Year 2, Prior experience teaching CHOICE	0.094	−0.063	0.251	1.17	0.30	−0.20	0.79
Year 2, Received CHOICE training	0.807	0.017	1.597	2.00	0.50	0.08	0.92
Baseline Organizational Support for Evidence-based Practices[a]	0.413	−0.459	1.285	χ^2 (2, n = 11) = 2.34, p = .311	0.28	−0.29	0.85
Baseline GTO knowledge [a]	0.125	−0.159	0.409		0.26	−0.32	0.84
CFIR (in Year 2)	0.640[*b]	0.411	0.869	5.48	0.78	0.46	1.10

Note: GTO site N = 13. Tabled values are causal coefficient estimates except where noted, with confidence intervals, test statistics, and standardized regression coefficients.

[a]Predictors tested simultaneously; reported test statistic is Wald χ^2 with 2 degrees of freedom.

[b]Not established as a causal estimate; path is potentially confounded within the framework of the graphical causal model.

[*]Statistically significant after false discovery rate correction, adjusted ps < .05.

Table 5.

Unstandardized and standardized causal coefficient estimates with confidence intervals for Year 3, GTO performance predicting three different measures of CHOICE fidelity in Year 3

Predicted by Year 3, GTO Performance	Causal coefficient (b)	Lower 95% CI	Upper 95% CI	z	Standardized coefficient (β)	Lower 95% CI	Upper 95% Ci
Year 3, MITI Global rating scores	0.16 *	0.07	0.26	3.50	0.68	0.29	1.07
Year 3, Evocation	0.76	−0.27	1.79	0.53	0.36	−0.22	0.94
Year 3, Adherence to CHOICE	1.73	−4.19	7.65	0.57	0.27	−0.58	1.12

Note: GTO site $N = 12$; CHOICE adherence rating $N = 264$. Table values are causal coefficient estimates with confidence intervals, test statistics, and standardized regression coefficients.

*
Statistically significant after false discovery rate correction, adjusted ps < .05.

Table 6.

Treatment differences in GTO and CHOICE outcomes in Year 3

	Regression coefficient (*b*)	Standard Error	Lower 95% CI	Upper 95% CI	*t* (22)	Cohen's *d*	Lower 95% CI	Upper 95% CI
Evocation	−0.325	0.324	−0.998	0.347	−1.00	−0.41	−1.22	0.41
% Complex Reflections	0.079	0.067	−0.060	0.220	1.18	0.48	−0.34	1.29
% Open Questions	0.111	0.055	0.000	0.230	2.01	0.77	−0.07	1.60
Reflection Question Ratio	0.179	0.098	−0.020	0.380	1.84	0.72	−0.12	1.54
% MI Adherent	−0.074	0.073	−0.230	0.080	−1.02	−0.42	−1.22	0.40
	Logistic coefficient				*t* (20)	Odds Ratio	Lower 95% CI	Upper 95% CI
Session-level adherence	−0.34				−0.50	0.71	0.17	2.98
	Frequency/ Observations (Percentage) GTO	Frequency/ Observations (Percentage) Control			Fisher's Exact Test p	Phi Coefficient		
Sessions administered (5 vs <5 sessions)[*]	11/13 (85%)	10/11 (91%)			1.00	−0.09		
Year 3, GTO performance								
Step 7	10/10 (100%)	7/10 (70%)			0.210	0.42		
Step 8[*]	9/9 (100%)	1/13 (8%)			<.001	0.91		
Step 9	11/11 (100%)	8/13 (62%)			0.041	0.47		
Step 10	8/13 (62%)	5/13 (38%)			0.434	0.23		

Note: Year 3 site *N* = 28; CHOICE Adherence rating *N* = 394. Upper table values are regression coefficient estimates with confidence intervals, test statistics, and effect size estimates. Lower table values are descriptive frequencies with exact test *p* values and effect size estimates.

[*] Statistically significant after false discovery rate correction, adjusted *p*s < .05.