Implementing a Measurement Feedback System: A Tale of Two Sites

Leonard Bickman[1], **Susan R. Douglas**[1], **Ana Regina Vides De Andrade**[1], **Michele Tomlinson**[2], **Alissa Gleacher**[3], **Serene Olin**[3], and **Kimberly Hoagwood**[3]

Leonard Bickman: leonard.bickman@vanderbilt.edu

[1]Vanderbilt University, Nashville, USA

[2]Hope Network, Grand Rapids, MI, USA

[3]New York University, New York, USA

Abstract

A randomized experiment was conducted in two outpatient clinics evaluating a measurement feedback system called contextualized feedback systems. The clinicians of 257 Youth 11–18 received feedback on progress in mental health symptoms and functioning either every 6 months or as soon as the youth's, clinician's or caregiver's data were entered into the system. The ITT analysis showed that only one of the two participating clinics (Clinic R) had an enhanced outcome because of feedback, and only for the clinicians' ratings of youth symptom severity on the SFSS. A dose–response effect was found only for Clinic R for both the client and clinician ratings. Implementation analyses showed that Clinic R had better implementation of the feedback intervention. Clinicians' questionnaire completion rate and feedback viewing at Clinic R were 50 % higher than clinicians at Clinic U. The discussion focused on the differences in implementation at each site and how these differences may have contributed to the different outcomes of the experiment.

Keywords

Measurement feedback system (MFS); Contextualized feedback systems (CFS); Youth mental health services outcomes; Peabody treatment progress battery (PTPB); Symptoms and functioning severity scale (SFSS); Implementation analysis; Routine outcome measurement (ROM)

Introduction

The dissemination and implementation of quality improvement tools, such as measurement feedback systems (MFSs), has been advocated as a means of improving service delivery and outcomes for children (American Psychological Association Presidential Task Force on Evidence-Based Practice for Children and Adolescents 2006; Bickman et al. 2011; Jensen-Doss and Hawley 2010; New Freedom Commission on Mental Health 2003; Sapyta et al.

2005). It has been suggested that MFS's (Bickman 2008) are a prime quality improvement tool because they provide ongoing data and feedback to clinicians and families about treatment progress. Research on the ability of MFSs to improve the quality of mental health services is continuing to grow as states, counties, and other entities expand their efforts to provide evidence-based services to youth and families (Bruns et al. 2008).

MFSs have a positive impact on outcomes in different subspecialties of medicine (Duncan and Pozehl 2000; Goebel 1997; Goodman et al. 2013; Holmboe, et al. 1998; Leshan et al. 1997; Mazonson et al. 1996; Robinson et al. 1996; Rokstad et al. 1995; Tabenkin et al. 1995), education (Arco, 1997; Arco 2008; Furman et al. 1992; Mortenson and Witt 1998; Rose and Church 1998; Tuckman and Yates 1980), and mental health (Bickman et al. 2011; Chorpita et al. 2008, 2011; Howe 1996; Lambert 2001; Lambert et al. 2001, 2005; Mazonson et al. 1996). While quality improvement tools, such as MFSs (Cebul 2008), have been successfully applied for several decades (Kluger and Denisi 1996; Rose and Church 1998), their use is not widespread within children's mental health services.

Recently,Bickman et al. (2011) investigated a MFS, called Contextualized Feedback and Intervention and Treatment (CFIT) that provided feedback on mental health progress and therapy process variables, specifically examining the impact of routine clinician feedback on youth outcomes. Using a randomized-cluster design they found that youth whose clinicians received session-by-session feedback improved more quickly than those where feedback was provided every 90 days. Moreover, a dose–response curve indicated stronger effects when the clinician accessed more of the available reports. The use of a MFS improved the rate of clinical progress.

This paper deals with the second generation of CFIT now called Contextualized Feedback Systems (CFS™; Bickman et al. 2011, 2012). The name change was due to a trademark issue. The theory underlying CFS remained the same (Riemer and Bickman 2011; Riemer et al. 2005) but a new software product was developed to update the software platform, incorporate changes based on user feedback, and make revisions based on accumulating knowledge of barriers and facilitators to MFS implementation. For example, the software was modified to track and provide feedback on measure completion rates automatically. This change, made in response to user requests, made it easier for clinicians to assess client and caregiver participation as well as what the clinicians needed to complete. Supervisors also could use the information to support clinicians in adhering to the intervention (e.g., completion of measures was necessary for feedback to occur and viewing of feedback was necessary to influence clinician behavior) by providing performance expectations and accountability measurement, both factors found to promote successful implementation (e.g., Fixsen et al. 2005).

Observational studies of MFSs indicate that the challenges in implementing MFSs can be substantial, with specific issues related to the introduction of technology in mental health care; the integration of quantitative feedback data with the typical person-centered, non-numerical values of clinicians (Bickman et al. 2000); concerns about the use of feedback data for performance evaluation; and practical considerations like time burden and resources (Boswell et al. 2015; Duncan and Murray 2012; Fitzpatrick 2012; Martin et al. 2011;

Wolpert et al. 2014). In addition, mental health clinicians, similar to other health care providers, have been found to overestimate their own abilities to assess treatment progress and outcomes based solely on clinical judgment (e.g., Hannan et al. 2005; Hatfield et al. 2009; Hawley and Weisz 2003; Love et al. 2007), which contributes to lower utilization of outcome management systems like a MFS (Boswell et al. 2015).

This current study focuses on the effects of CFS on outcomes and secondarily tries to make sense of the implementation issues that arose in the two sites studied. This paper reports on the quantitative effects of feedback on youth mental health progress. In addition it provides some observations on the implementation of the MFS at two sites. A companion paper (Gleacher et al. this issue) applies a qualitative methodology and a different theoretical framework to examine implementation.

As described below in the methods, the types (e.g., manuals, onsite training, ongoing coaching), timing, and content of implementation strategies were the same for the two sites. However, consistent with high quality implementation strategies (e.g., Fixsen et al. 2005), support was tailored to meet the needs and resources of each site. We use the conceptual framework provided by Powell and his colleagues, (Powell et al. 2012) to describe how different implementation strategies emerged from the two sites. These authors define strategies as the systematic process used to integrate evidence-based interventions into usual care. They compiled a list of implementation strategies after reviewing 205 sources published between 1995 and 2011. Their list includes 68 strategies and definitions, which are categorized into six key implementation processes: planning, educating, financing, restructuring, managing quality, and attending to the policy context. Using their list of strategies, we compare the two sites on which strategies, as actually implemented, were the same or different.

Methods

Design and Procedures

The New York State's Office of Mental Health in 2005 began funding for an Evidence-Based Treatment Dissemination Center (EBTDC) to train clinicians across the state on a range of evidence-based clinical practices for children (Gleacher et al. 2011). This project was built on EBTDC efforts to improve the quality of services, and was designed to assess the feasibility and impact of integrating CFS in community mental health clinics (CMHCs). Thirty mental health agencies, which had at least five clinicians previously trained in evidence-based practices (EBPs), were invited to participate in the study. Of the 11 agencies that applied to participate, two were selected on the basis of their experience and success in implementing other state initiatives. Those two agencies had two clinics each, and thus a total of four clinics began the study. Due to concerns about burden on clinicians resulting from statewide restructuring of clinic services and MFS implementation occurring simultaneously (see Gleacher et al. this issue for more information), one agency withdrew at about 9 months after selection and 3 months after the delayed roll out (described further below). Thus, CFS was implemented for 2 years in two geographically separated clinics at a single agency.

The participating agency provided outpatient mental health services to youth and their families. Mental health services were designated by the agency as evidenced-based practices (EBPs). The two clinics differed in location, with one urban (for convenience called Clinic U) and one rural (called Clinic R). All clients new to services aged 11–18 years were eligible to participate in the study. Clients and their families were informed by agency staff of the quality improvement initiative at intake and given the opportunity to decline participation. All clinicians and their supervisors were mandated to participate in the study as part of the agency's quality improvement efforts.

There were two experimental groups at each site: the intervention group, where clients had session-by-session feedback reports generated for clinician review and the control group, where clients had feedback reports generated every 6 months. CFS feedback included data on treatment progress (e.g. symptoms and functioning) and process (e.g. therapeutic alliance) for that client as soon as measures were entered in the system. All cases involved in the study were asked to complete the same measures at the same frequency. Random assignment was at the client-level, meaning that one clinician could have both feedback and control cases. Clients associated with a given clinician were randomized in blocks of four clients by the evaluation team. Immediately after the initial intake appointment was completed and relevant intake measures were submitted in CFS, an automated process within the CFS application randomly assigned clients to a feedback or a control condition. The New York University and Vanderbilt University Institutional Review Boards approved all procedures with a waiver of informed consent.

Feedback Intervention and Support

CFS is a web-based application that provided computerized feedback reports to agency personnel (e.g., directors, supervisors, clinicians) based on a battery of psychometrically sound and clinically useful measures (Bickman and Athay 2012). CFS is a theory-based intervention intended to promote overall practice improvement through frequent and comprehensive assessments that produce feedback that can change clinician behavior (Riemer and Bickman 2011; Riemer et al. 2005).

The study protocol recommended CFS measures be administered to clinicians, clients, and a caregiver (if present) within the session during the closing 5–10 min. Based on personal preference and time constraints, clients and their caregivers could also complete measures immediately after a session, with clinicians attempting to complete their relevant measures within a few days of the session. All measures could be completed electronically using computers or tablet devices or by paper-and-pencil and manually data entered. Feedback reports were available to be viewed or printed by clinicians and their supervisors as soon as data were entered in the web-based application.

Because a MFS represents an additional source of innovative, technologically based information, and a deviation from the typical means of assessing progress in CMHCs, the training and consultation followed a set of steps directed by the developers of CFS. The CFS training and consultation model drew upon previous experience (Bickman et al. 2011; Douglas et al. 2014) and multiple theories of learning in addition to research on implementation of health information technology (Wager et al. 2009). Thus, a mix of

'training best practices' formed the structure of the CFS training and consultation including: (1) the development of site-based 'communities of systematic inquiry', or local site-based workgroups, consisting of key stakeholders from direct line staff to leadership staff within an agency; (2) ongoing consultation and coaching from CFS consultants throughout implementation, decreasing in intensity over time; and, (3) additional training and support for key staff identified as 'master partners', or super-users of the CFS system. Each of these strategies was intended to build local ownership and investment to enhance implementation fidelity and intervention sustainability over time. These goals were accomplished through three phases of training and consultation conducted over a period of about two years. There were three phases to the implementation: (1) pre-implementation contextualization phase: This phase focused on providing guidance to sites on how to prepare for CFS integration into existing clinic procedures and protocols via collaborative workgroups; (2) action learning phase: Staff were oriented to CFS operations through "how-to" webinars, applied exercises using sample cases, and reference manuals; and, (3) post-implementation coaching phase: Select staff engaged in regularly scheduled webinars and teleconference calls to support ongoing clinical and operational utilization of CFS (including active review of aggregated implementation data), with additional support (e.g. site visits) available as needed.

Support activities were contextualized based on agency/ clinic needs and research team resources and designed to meet changing needs over time. At the start of the project, Vanderbilt University (VU) staff conducted all phase one and two activities with participation by Columbia University (CU) (investigators later moved to New York University—NYU) project staff as learner-observers. Due to a six-month postponement in the availability of the CFS application as a result of software development delays, the phase one pre-implementation and contextualization activities were extended to span the 6 months period prior to roll out. Phase two, the active learning phase, occurred as planned after phase one with CFS available and ready to be used with clients.[1] Phase three support, post-implementation coaching, was originally intended to be provided directly to all sites by CU project staff, with secondary support and consultation available from VU staff. However, due to resource limitations, the phase three support activities started out with CU project staff providing support to one agency and VU staff to the other agency. After the first agency dropped out, phase three activities were modified to best fit needs at the time with CU project staff providing in-person consultation on a monthly basis to the urban clinic (Clinic U) and an agency-based higher-level administrator and her assistant providing in-person consultation at least monthly to the rural clinic (Clinic R). Vanderbilt staff provided ongoing consultation directly to the CU project staff and the higher-level administrator and her assistant. Notably, the agency issued a mandate about midway through the project for both sites requiring the use of CFS for all eligible cases with CFS implementation data to be used in staff performance evaluations.

1Note that this initial version was considered a beta version with software bugs identified and corrected throughout about the first year.

Treatment Measures

Treatment progress and process measures used by CFS in the current study were from the Peabody Treatment Progress Battery (PTPB; Bickman et al. 2010). The PTPB includes eleven measures completed by youth, caregivers, and/or clinicians that assess clinically-relevant constructs such as symptom severity, therapeutic alliance, life satisfaction, motivation for treatment, hope, treatment expectations, caregiver strain, and service satisfaction. All PTPB measures have undergone multiple rounds of comprehensive psychometric testing and demonstrate evidence of reliability and validity in this population for all respondent versions (Bickman et al. 2010). All eleven measures were completed as part of CFS and resulting data used to create clinical feedback reports. Symptom severity was the outcome variable of focus in the current analyses.

The Symptoms and Functioning Severity Scale (SFSS: Bickman et al. 2010) is a measure of symptom severity with parallel forms completed by the clinician, caregiver and youth. Composed of 26 five-point Likert-type items (27 for the clinician version), it yields a total score of global symptom severity as well as subscale scores for internalizing and externalizing symptom severity. In the current paper, the global severity score was used in analyses. The SFSS has demonstrated sound psychometric qualities for all three respondent forms including internal consistency (range $\alpha = 0.93$–0.94), test–retest reliability (range $r = 0.68$–0.87), construct validity, and convergent and discriminant validity. See Athay et al. (2012) for more information about the SFSS.

Demographic and background information (e.g., race and ethnicity) for clients and their caregivers were routinely collected as part of the initial intake measurement package. Clinicians also completed an initial background questionnaire during the phase two active learning activities.

Results

Client, Clinician and Session Characteristics

Client Characteristics—The sample used in these analyses consisted of youth receiving mental health treatment who were at least 11 years old and had completed at least one SFSS measure. SFSS data were scheduled to be collected after each session from three respondents: the youth, caregiver, and clinician. A total of 257 youth (hereafter the "analytical sample") had 2698 clinical sessions with at least one completed SFSS measure present (1258 in Clinic U and 1440 sessions in Clinic R). The analytical sample and sessions were generally evenly spread across sites and treatment groups (Table 1).

Table 2 shows baseline characteristics for youth and caregivers between conditions within site. As expected, randomization was successful and client characteristics such as ethnicity, age, gender, and baseline functioning level were similar between groups within site (all bootstrapped $p > .05$). More information comparing sites can be found in the Gleacher et al. (this issue) paper.

Clinician Characteristics—Twenty-one clinicians provided information in the CFS data system, 13 from Clinic U and eight from Clinic R. Regardless of site, most of the clinicians

were female (75 %), white (62 %), non-Hispanic (90 %), and between 26 and 30 years old (43 %). Nearly all (90 %) clinicians had attained a master's (80 %) or doctoral (10 %) degree, typically in Social Work (47 %). A third (33 %) of clinicians reported subscribing to a cognitive-behavioral therapeutic orientation. However, most (48 %) reported some other unspecified orientation or that they had no particular therapeutic orientation. About half (52 %) of the clinicians were licensed to practice in the state in which they currently worked. Most (47 %) had less than 1 year of experience at their current work place. Relatively few clinicians had more than 5 years of experience anywhere— either in their current workplace (10 %) or elsewhere (19 %). About a fifth (19 %) of the clinicians had no experience in providing services for children or youth either in their current work place or in any other place before they used CFS. The Gleacher et al. paper (this issue) provides further information on the differences between sites.

Episodes—Treatment for mental health was often episodic, where clients received regular treatment for a period of time and discontinued it for various reasons (e.g., they improved and no longer required treatment; they stopped coming for logistical reasons; etc.). However, clients could start up treatment again at a later date. This was called a second treatment episode. In this study, an episode was determined to have ended if no clinical sessions were held for 60 days. If two consecutive sessions were more than 60 days apart, the later session reflected a new episode. Our analytical sample included only one single treatment episode per youth (keeping the first episode unless the first episode only had two or fewer data points and the second episode had three or more data points). Descriptive data for clients' treatment episodes by site can be found in Table 3. No significant differences in treatment episode characteristics were found between conditions within site.

Session Attendance—This study had access to billing information on study participants. Based on billing information, we could determine whether a clinical session was held as scheduled or whether the session was not held and why. Sessions that were not held either were cancelled or were "No-Show" appointments where the youth and/or caregiver did not show up for their scheduled appointment. Table 3 describes the clinical sessions held within one episode and the clinical sessions not held per condition within each site. For example, in the control group in Clinic U, youth had an average of 10.39 clinical sessions within one treatment episode with an average of 3.08 other scheduled clinical sessions that were not held.

According to the CFS intervention protocol, youth, caregivers and clinicians should have completed an SFSS measure every time a clinical session was held. Therefore, the total number of sessions that were held (according to billing records) was used to calculate the SFSS completion rates. Table 3 shows the SFSS completion rates per respondent (youth, caregiver, and clinician). For instance, a 100 % youth SFSS completion rate would mean that at every clinical session held, the youth always filled out the SFSS measure. In Clinic U, youth's and clinicians' SFSS questionnaire completion rates were on average 60 and 45 % respectively. In Clinic R, youth' and clinicians' completion rates were very similar at 65 and 69 % respectively. There were no significant differences in questionnaire completion rates between conditions within each site. Caregivers' average questionnaire completion rates

were lower than the youth or clinician in both conditions and sites. It is important to note that the caregivers' questionnaire completion rates are an underestimate of the true completion rate as these rates were calculated based on the total number of treatment sessions *held for each youth*. However, the youth's caregiver might not have been present at each clinical session held and therefore would not be expected to complete the functioning questionnaire. Although CFS was designed to collect information from the caregiver regardless of whether they were present in the session, we believe that questionnaires may not have been given to caregivers if they were present in the clinic but not in the session. Therefore, these lower completion rates do not necessarily reflect refusal or incomplete questionnaire completion on the part of the caregiver; they may, in fact, reflect lack of adherence to the CFS protocol or lower attendance of caregivers since we are not able to determine if a caregiver was present during the session.

Mode of SFSS Completion—There were two ways that data were entered into CFS: manually or electronically. Respondents could complete paper–pencil SFSS forms, and then their answers were manually entered into the computerized system, or they could complete the SFSS measures online (e.g., on a desktop computer, tablet) directly into the system. Both methods were available and used at both sites but site differences in mode of completion emerged. At Clinic U, the majority of youth and caregiver SFSS's were completed by hand then manually entered (61 and 78 % respectively) whereas at Clinic R, the minority of youth and caregiver SFSS's were completed on paper (14 and 40 % respectively). In both sites, nearly all (97 %) clinician SFSS data were entered electronically. Despite site differences, no significant differences were seen between treatment and control conditions within site (Table 4).

Treatment Effect—Intent to Treat Analyses

This study aimed to evaluate the effects of providing feedback on clients' progress to clinicians. The first hypothesis examined whether systematic session-by-session feedback on clients' progress provided to clinicians improves clients' symptoms and functioning outcomes more than similar clients whose clinicians did not receive any feedback. A second hypothesis examined whether there was a dose effect on youth outcomes for those clients in the feedback condition. Each site was treated independently in the analyses because of the differences in implementation that are discussed later in this paper and in the accompanying implementation-focused paper that is part of this special issue (Gleacher et al. this issue).

Clinic U—Intent to treat (ITT) results for Clinic U are shown in Table 5. The average youth-rated symptom severity was 2.21 at the beginning of the episode and significantly decreased (i.e. improved) over the course of treatment (β slope intercept = −0.02, $p < .0001$). On average, caregiver- rated symptom severity was 2.53 at the beginning of the episode and significantly decreased over the course of treatment (β slope intercept = −0.01, $p = .049$). The average clinician-rated SFSS was 2.26 but did not change significantly over the treatment episode. There were no significant feedback effects for Clinic U.

Clinic R—ITT results for Clinic R are found in Table 6. Average youth-rated symptom severity was 2.23 at the beginning of the episode and significantly decreased (improved)

over the course of treatment (β slope intercept = −0.02, $p < .0001$). The average caregiver-rated symptom severity was 2.29 at the beginning of the episode and significantly decreased over the course of treatment (β slope intercept = −0.01, $p = .001$). On average, the clinician-rated SFSS was 2.18 at the beginning of the episode and symptom severity significantly decreased through treatment for youth in the feedback group only (β feedback* slope = −0.01, $p = .045$). Thus, the feedback effects were seen only in the clinician ratings.

The Implementation Index (Dose of Feedback)

Next, analyses were conducted to investigate whether there was a dose effect on youth outcomes for those in the feedback condition. To answer whether youth's improvement increased with the amount of feedback the clinician received (dose), we calculated a composite *Implementation Index*. The implementation index captured the potential amount of the feedback intervention received by the clinician. The implementation index incorporated two dimensions: (1) each respondent's questionnaire completion rate and (2) the clinicians' feedback viewing rate. The CFS system indicated whether the feedback was accessed but not if the clinician actually read the feedback or used it in providing services. Thus, it is a rather crude measure of implementation and at best shows only the potential for applying the feedback to each case. Moreover, it does not show and if and how the clinician used the feedback information in treatment. The full implementation, using this index of CFS use, requires at a minimum both the completion of the SFSS questionnaires by each respondent *and* the clinician viewing (accessing) every resulting feedback report created automatically after data entry from each session. Both of these actions must happen to at least some reasonable degree for successful implementation to occur. In other words, the intervention being evaluated (i.e., the use of feedback in clinical practice) is not actually delivered if questionnaires are completed but feedback is never viewed, and feedback cannot be created without the data actually gathered from completed questionnaires.

Because both actions, questionnaire completion and feedback viewing, must occur for clients to receive the intended intervention, the estimated Implementation Index includes both key components: questionnaire completion rates and clinician's feedback viewing. To calculate the Implementation Index, the questionnaire completion rates for the SFSS from the three respondents (youth, caregiver, and clinician) were averaged over all sessions for each youth. Then, the clinician feedback viewing rates for each SFSS measure from the three respondents were averaged for each youth. These two averaged scores were then *multiplied* together and divided by 100 so the implementation index ranged from 0 to 100. In effect, the multiplication 'weighs' the feedback viewing; viewing ensures implementation, but questionnaire completion is the input needed to create the feedback report. Therefore, failure to fill out the questionnaires (questionnaire response) and failure to view the feedback results in CFS implementation failure (i.e., the conditions that define the control group). Table 7 provides some specific implementation index scores, examples of how they are achieved, and their interpretation.

The average questionnaire completion and feedback viewing per SFSS respondent and the resulting implementation index by site are found in Table 8. The average implementation index for Clinic R was 7 %points higher than Clinic U. The major differences between

clinics are in the clinicians' behavior. Clinicians' SFSS implementation index of Clinic R were 26 % points higher than clinicians of Clinic U.

Figure 1 depicts the distribution of the clients' implementation index for both sites. In Clinic U (left panel), the distribution was skewed to the right, indicating the mode was zero (no implementation), 41 % of youth in the feedback group (23 out of 56 youth) received 0 to 10 % of the intervention, and only 18 % (10 out of the 56 youth) received 50 or more percent of the intervention. On the other hand, in Clinic R the distribution is a more symmetrical, with a mean and median of 34 %, a mode of 50 %, 30 % of youth in the feedback group (22 out of 75 youth) received 50 % or more of the intervention and 21 % received 0 to 10 %. In summary the Implementation Index shows that Clinic U had very poor implementation (a mode of zero) while in Clinic R a mode of 50 % implementation was obtained.

Dose–Effect Results

Clinic U—Table 9 shows the HLM dose effect results for Clinic U. These analyses are similar to the ITT analysis except that they are based only on the feedback group and the implementation index is added as a covariate instead of treatment group. As with the previous ITT analysis, the dose analyses are estimated separately by site. For Clinic U, regardless of respondent, there was no dose effect. The implementation index by slope-estimated coefficients (index* slope) were not significant (all $p > 0.05$). Consistent with the ITT analyses youth improved during the treatment phase regardless of dose of the intervention when outcome was assessed by the youth (β slope intercept = −0.02, p < .001) and by the clinician (β slope intercept = −0.01, $p = 0.006$).

Clinic R—Table 10 shows the Clinic R dose effect results. There is a statistically significant dose–effect when the outcome is rated by the youth and the clinician (both implement*slope estimated coefficients $p < .05$). For the youth SFSS, when the implementation index is zero (i.e., the youth's clinician did not receive any CFS intervention) the β Slope Intercept was 0.018, $p = 0.007$ indicating that clients' functioning symptoms with no CFS intervention got worse over the course of the treatment period. However, youth symptoms improved significantly faster with CFS intervention (β implement*slope = −0.001 for each additional implementation unit, $p < .001$). According to the clinician SFSS model, clients with zero implementation did not improve (β slope intercept = 0.004, $p = 0.478$) while youth improved faster the better the implementation of the CFS intervention (β implementation*slope = −0.0003 per each additional implementation unit, $p = 0.021$). There was no dose–effect for caregiver ratings.

Effect Size Estimates—Because of the time-varying nature of the dose–effect, the effect size changed by time in treatment and level of the CFS implementation. Table 11 shows the estimated effect size of this dose–effect relationship at the 8th, 10th and 15th week during the treatment phases when the implementation index is 0, 25, 50, and 75. For both the youth and clinician SFSS models, the dose–effect grows stronger throughout treatment and is the strongest when implementation was high.

Summary

The ITT analysis showed that only one of the two participating clinics (Clinic R) had an enhanced outcome because of feedback, and only for the clinicians' ratings of youth symptom severity on the SFSS. A dose–response effect was found only for Clinic R for the client and clinician ratings. Based on these data, if the implementation index was 75 % we estimated that effect sizes of between 0.7 and 0.8 would be found after 10 weeks of treatment. Detailed implementation analyses showed that Clinic R had better implementation of the feedback intervention. The superior implementation was based primarily on the clinician's behavior. Clinicians' SFSS questionnaire completion rate and feedback viewing at Clinic R were 50 % higher than clinicians at Clinic U. Mode of data entry was also different between clinics with clinic R caregivers and clients more frequently using the computerized system instead of paper and pencil. Computerized data entry in both clinics by clinicians was equally high (over 95 %).

Discussion

The results of this study provide support for the efficacy of feedback to clinicians in improving client outcomes. According to the criteria for evaluating evidence-based treatments (e.g., Southam-Gerow and Prinstein 2014), CFS meets the definition of a Level 2 intervention as a "probably efficacious treatment." With two randomized control trials in two independent research settings, the only criterion missing to achieve a Level 1 designation as a "wellestablished treatment," is replication by two independent research teams.

In a randomized design Clinic R showed that youth improved faster when the clinician was provided feedback on the clinician's measure of severity. Moreover, there was a dose–response effect for the feedback group in Clinic R for both the clinician and the youth's outcome measures. There was no feedback effect on the caregivers' or youth's measures. It is plausible that the lack of effect on the caregivers' measure may be due to the lower power of that sample. For this study, the randomization at the client level was an improvement from a research methodology perspective; however, it raised the possibility that clinicians may experience difficulty by managing cases both with and without feedback available. Notably, the rates of feedback viewing were similar to a previous study (Bickman et al. 2011) using CFS where randomization occurred at the site level. Anecdotally, it was noted that some clinicians who valued feedback would ask when they could get feedback reports for all their clients. After the study was completed, the agency leadership decided independently to expand the use of a MFS across the organization.

There were no effects of feedback found in Clinic U, which had substantially lower implementation index scores. The discussion now turns to possible reasons why Clinic U had poorer implementation than Clinic R. While there were no significant differences between the sites in the mental health and demographic characteristics of the youth and clinicians, the clinic environments were different. (For additional discussion of implementation barriers and facilitators, including findings from semi-structured staff interviews, see the accompanying article by Gleacher et al. this issue).

As noted in the introduction, Powell et al. (2012) described implementation strategies as falling into the following categories: (1) Planning strategies (n = 17) include such activities as conducting actions like a local needs assessment and building coalitions; (2) Education strategies (n = 16) include developing educational materials and ongoing training; (3) Finance strategies (n = 9) include financial incentives and reducing fees; (4) Restructuring strategies (n = 7) include revising professional roles and changing equipment; (5) Quality management strategies (n = 16) include quality monitoring and use of data experts; and (6) attending to policy context which includes changing accreditation and liability laws.

The authors examined each of the 68 strategies and categorized them into same or different depending on clinic. We did not use any of the restructure, finance or policy context implementation strategies at either site. The following strategies were used in a similar fashion in both sites; assess for readiness and identify barriers, mandate change,[2] develop academic partnerships, conduct ongoing training, use train-the-trainer strategies, facilitate relay of clinical data to providers; and change records systems. We believe that two major implementation differences between the clinics related to the following two strategies: (1) how the feedback measures were administered, (2) how ongoing consultation and leadership support was provided, including the provision of clinical supervision.

Although the CFS software was the same in both sites, the way the feedback was implemented was different. Clinic U did not use computers or the provided tablets for youth or caregivers to answer the questionnaires at the end of the session. This resulted in delay in the clinicians receiving the feedback since the paper forms had to be hand entered. There were no additional resources provided for data entry in either site so this process took longer. The additional delay could have weakened the effectiveness of the feedback by decreasing the value of the feedback data as less relevant to the current clinical picture (i.e., it could be several weeks old). Possible reasons for less frequent use of computerized data entry at Clinic U are provided in the accompanying paper by Gleacher et al. (this issue).

At Clinic R, an internal senior clinic administrator and her assistant provided the ongoing consultation, clinical supervision and technical support in the use of CFS. The senior administrator was located in the central office, in close proximity to Clinic R, and provided day-to-day oversight of CFS implementation. At Clinic U, CU project staff provided the consultation. While the project staff may have been as technically competent as the internal senior agency staff it is plausible to assume that the senior internal staff had greater influence on the behavior of the clinicians. Clinic R having greater day-to-day involvement of senior leadership staff at Clinic U may have generated greater engagement of all staff, even though staff at both clinics were required to complete CFS measures as part of performance evaluation. Another possibility is that having in-house staff support resulted in accessibility to supports that facilitated clinician implementation behavior; Clinic U had to rely on external staff (project staff) to provide in-person support on a scheduled basis.

---

[2]As noted the methods section, participation in the quality improvement initiative was required of agency staff. In addition, the agency issued a mandate about midway through the project for both sites requiring the use of CFS for all eligible cases with CFS implementation data to be used in staff performance evaluations.

Our reporting of implementation differences was placed in the discussion section of this paper because of its speculative nature. We do not believe that the observational and post hoc nature of the data has the same validity as our outcome analyses. They are provided here more as hypotheses worthy of further exploration.

What has not been discussed, but was alluded to in the method section, is our biggest implementation issue. The software needed to operate CFS was complex and was constantly under development. Although this was the second version of the software, a different programming company was responsible for development. The story behind this development and one that took place later is both complex and tragic. There are two lessons here that are worth noting. First, it is difficult to fly a plane when you are building it (see http://www.youtube.com/watch?v=M3-hge6Bx-4w). Clearly it was not our intent to start the study without having a solid software package in hand. However, our start date was based on the time schedule provided by the software company and the research grant. What we were not sufficiently apprised of was the lack of reliability of such schedules in the software development field. While reliable and valid data are not readily available, the best estimates are that most software projects are more likely not to come in on time or on budget (Moløkken and Jørgensen 2003). This led to lesson two: Do not base the timing or budgets of projects on software that has not already been developed and fully tested. Technology is seductive for it makes things possible that we could not otherwise accomplish but its use entails significant risks.

The technological issues we encountered directly impacted implementation in several ways. First, in addition to negatively impacting implementation because of delays and development bugs, the technology was still so new that some barriers were not even known to exist until the clinicians encountered them. For example, we spent a month problem solving why CFS did not work properly at one site, only to find out that it had nothing to do with CFS (the issue was caused by internal security measures). Second, the impact of these barriers on training and support cannot be overstated—they justifiably decreased clinic and the research staff's faith in the reliability of CFS. In addition, often when clinic staff brought new barriers to our attention, we did not know if we should address it as an attitudinal issue (i.e., work to decrease expectations that the system itself was the source of the barrier) or whether there truly was a technological issue. Often there was nothing wrong with the system; it was a user error or a contextual factor that was the problem. Third, the CFS programming design team made mistakes along the way, perhaps being too quickly responsive to end user feedback on desired improvements to the first version. For example, CFS feedback in version 2 included a new element that provided an alert to the clinician of the overall number of 'red flags' that had yet to be viewed. It seemed like a good idea on the design board but failed in actual use because it occurred too frequently with insufficient information. Not all feedback interventions have had positive outcomes. In Kluger and DeNisi's (1996) original meta-analysis over a third of the interventions resulted in poorer performance. While CFS as a whole met many of the expectations for a successful feedback intervention (Noell and Gansle 2014), perhaps this element was too focused on metatask processes and took attention away from the primary task of attending to client progress.

Despite the significant and ultimately unsolvable challenges in software development and the formidable implementation roadblocks, this study's results show that feedback to clinicians, under conditions of 'good enough' fidelity, facilitates client improvement. As Baumberger (2014) and others state, with any intervention there is likely a 'tipping point at which treatment integrity and quality are sufficient to generate positive outcomes (p. 42).' The implementation index, which combined rates of questionnaire completion and feedback viewing, may be a useful tool in the quest to determine what levels of MFS implementation are sufficient to achieve desired outcomes. As can be seen from these study results, Clinic U's implementation was so low as to be labeled an implementation failure, and thus the lack of results was due more to poor implementation rather than to the effectiveness of the intervention itself. At Clinic R, it is notable that implementation was far from 100 %. Yet, an implementation index score of 34 % was enough to achieve some significant effects of feedback on youth outcomes. Future research should investigate further how best to report implementation indices not only in the discussion of research findings but also as reported to the MFS users (i.e., clinicians, supervisors, leadership staff). CFS reported rates of questionnaire completion and feedback viewing separately. It may be that the implementation index, as a single number, may have been as or more effective at promoting change in behavior. In addition, benchmarks for staff compliance were determined by agency managers' judgment on what was feasible rather than any data-informed approach. At a minimum, it is vital for MFS researchers to report implementation rates in any presentation of outcomes. Ideally, future research could examine the effect of varying performance expectations for implementation on client outcomes to better determine implementation benchmarks for quality review.

As of this writing CFS as a web-based application has been discontinued due to the continuing technological challenges and the expense in trying to meet these challenges. While this is an unfortunate result, the more that decade-long journey we have taken in the development and evaluation of CFIT, and later CFS, has helped establish evidence that a quality improvement tool like a MFS can advance youth mental health outcomes without regard to treatment modality or setting. We have also learned a great deal about implementation of MFSs. As a final coda, we would like to emphasize the wise words of Michael Porter, a Harvard Business School professor who was recently quoted as saying, "I think the big risk in any new technology is to believe the technology is the strategy (Useem 2014)." Our lesson learned here is that the heart of a MFS truly is the measures used and the utilization of the feedback to influence what happened in the very next session as well as being integrated into other clinical practices such as treatment planning and reviews, supervision, and program planning.

As both a technological intervention and a quality improvement tool, MFSs meet the definition of a 'disruptive innovation,' (Rotheram-Borus et al. 2012). We would add our voices to others in suggesting that all clinicians should utilize a MFS to inform practice (e.g., Weisz et al. 2014). As technology matures, more and more feedback systems will come to market. In order to inform implementation and further development of high quality MFSs, future research questions should include determining the ideal set of measures (single instrument or battery measuring treatment progress and/or process?); the frequency of administration and feedback viewing; how to set performance expectations for setting

implementation benchmarks; how to most effectively use the feedback in clinical practice; and, who should receive feedback (clinician or clinician and client?).

References

American Psychological Association (APA) Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. American Psychologist. 2006; 61(4):271–285. [PubMed: 16719673]

Arco L. Improving program outcome with process-based feedback. Journal of Organizational Behavior Management. 1997; 17(1):37–64.

Arco L. Feedback for improving staff training and performance in behavioral treatment programs. Behavioral Interventions. 2008; 23(1):39–64.

Athay MM, Riemer M, Bickman L. The Symptoms and Functioning Severity Scale (SFSS): Psychometric Evaluation and Differences of Youth, Caregiver, and Clinician Ratings over Time. Administration and Policy in Mental Health and Mental Health Services Research. 2012; 39(1–2): 13–29. [PubMed: 22407556]

Baumberger, BK. Understanding and promoting treatment integrity. In: Hagermoser Sanetti, LM.; Kratochwill, TR., editors. Treatment integrity: A foundation for evidence-based practice in applied psychology. Washington, DC: American Psychological Association; 2014. p. 35-54.

Bickman L. Why don't we have effective mental health services? Administration and Policy in Mental Health and Mental Health Services Research. 2008; 35:437–439. [PubMed: 18931904]

Bickman L, Athay MI. Youth Mental Health Measurement (special issue) Administration and Policy in Mental Health and Mental Health Services Research. 2012

Bickman, L.; Athay, MM.; Riemer, M.; Lambert, EW.; Kelley, SD.; Breda, C.; Tempesti, T.; Dew-Reeves, SE.; Brannan, AM.; Vides de Andrade, AR., editors. Manual of the Peabody Treatment Progress Battery. 2nd ed.. Nashville, TN: Vanderbilt University; 2010. [Electronic version]

Bickman L, Kelley DS, Athay M. The Technology of Measurement Feedback Systems. Couple and Family Psychology. 2012; 1(4):274–284. [PubMed: 24066274]

Bickman L, Kelley SD, Breda C, deAndrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. Psychiatric Services. 2011; 62(12):1423–1429. [PubMed: 22193788]

Bickman L, Rosof J, Salzer MS, Summerfelt WT, Noser K, Wilson SJ, Karver MS. What information do clinicians value for monitoring adolescent client progress and outcome? Professional Psychology: Research and Practice. 2000; 31(1):70–74.

Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. Psychotherapy research. 2015; 25(1):6–19. [PubMed: 23885809]

Bruns E, Hoagwood K, Hamilton J. State implementation of evidence-based practice for youths, Part I: Responses to the state of the evidence. Journal of the American Academy of Child & Adolescent Psychiatry. 2008; 47(4):369–373. [PubMed: 18356704]

Cebul RD. Using electronic medical records to measure and improve performance. Transactions of the American Clinical and Climatological Association. 2008; 119:65–75. discussion 75-76. [PubMed: 18596863]

Chorpita B, Bernstein A, Daleiden E. Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. Administration and Policy in Mental Health and Mental Health Services Research. 2008; 35:114–123. [PubMed: 17987376]

Chorpita BF, Bernstein AD, Daleiden EL. Empirically guided coordination of multiple evidence-based treatments: An illustration of relevance mapping in children's mental health services. Journal of Consulting and Clinical Psychology. 2011; 79:470–480. [PubMed: 21787050]

Douglas S, Button S, Casey SE. Implementing for sustainability: Promoting use of a measurement feedback system for innovation and quality improvement. Administration and Policy in Mental Health and Mental Health Services Research. 2014

Duncan EAS, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. BMC Health Services Research. 2012; 12:96. [PubMed: 22506982]

Duncan K, Pozehl B. Effects of performance feedback on patient pain outcomes. Clinical Nursing Research. 2000; 9(4):379–397. [PubMed: 11881696]

Fitzpatrick M. Blurring practice-research boundaries using progress monitoring: A personal introduction to this issue of Canadian psychology. Canadian Psychology/Psychologie Canadienne. 2012; 53(2):75–81.

Fixsen, DL.; Naoom, SF.; Blase, KA.; Friedman, RM.; Wallace, F. Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute National Implementation Research Network (FMHI Publication #231); 2005.

Furman CE, Adamek MS, Furman AG. The use of an auditory device to transmit feedback to student therapists. Journal of Music Therapy. 1992; 29(1):40.

Gleacher AA, Nadeem E, Moy AJ, Whited AL, Albano A, Radigan M, Wang Rui, Chassman J, Myrhol-Clarke B, Hoagwood KE. Statewide CBT training for clinicians and supervisors treating youth: The New York State Evidence Based Treatment Dissemination Center. Journal of Emotional and Behavioral Disorders. 2011; 19(3):182–192.

Goebel LJ. A peer review feedback method of promoting compliance with preventive care guidelines in a resident ambulatory care clinic. The Joint Commission Journal on Quality Improvement. 1997; 23(4):196–202. [PubMed: 9142611]

Goodman JD, McKay JR, DePhilippis D. Progress monitoring in mental health and addiction treatment: A means of improving care. Professional Psychology: Research and Practice. 2013; 44(4):231–246.

Hannan C, Lambert M, Harmon C, Nielson S, Smart D, Shimokawa K, et al. A lab test and algorithms for identifying clients at risk for treatment failure. Journal of Clinical Psychology. 2005; 61(2): 155–163. [PubMed: 15609357]

Hatfield D, McCullough L, Frantz SHB, Krieger K. Do we Know When our Clients Get Worse? An Investigation of Therapists' Ability to Detect Negative Client Change. Clinical Psychology and Psychotherapy. 2009

Hawley KM, Weisz JR. Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. Journal of Consulting and Clinical Psychology. 2003; 71(1):62–70. [PubMed: 12602426]

Holmboe E, Scranton R, Sumption K, Hawkins R. Effect of medical record audit and feedback on residents' compliance with preventive health care guidelines. Academic Medicine. 1998; 73(8): 901–903. [PubMed: 9736851]

Howe A. Detecting psychological distress: can general practitioners improve their own performance? British Journal of General Practice. 1996; 46(408):407–410. [PubMed: 8776911]

Jensen-Doss A, Hawley KM. Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. Journal of Clinical Child and Adolescent Psychology. 2010; 39:885–896. [PubMed: 21058134]

Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin. 1996; 119(2):254–284.

Lambert MJ. Psychotherapy outcome and quality improvement: introduction to the special section on patient-focused research. Journal of Consulting and Clinical Psychology. 2001; 69(2):147–149. [PubMed: 11393592]

Lambert MJ, Hansen NB, Finch AE. Patient-focused research: using patient outcome data to enhance treatment effects. Journal of Consulting and Clinical Psychology. 2001; 69(2):159–172. [PubMed: 11393594]

Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. Journal of Clinical Psychology. 2005; 61(2):165–174. Review. [PubMed: 15609358]

Leshan LA, Fitzsimmons M, Marbella A, Gottlieb M. Increasing clinical prevention efforts in a family practice residency program through CQI methods. The Joint Commission Journal on Quality Improvement. 1997; 23(7):391–400. [PubMed: 9257178]

Love SM, Koob JJ, Hill LE. Meeting the challenges of evidence-based practice: Can mental health therapists evaluate their practice? Brief Treatment and Crisis Intervention. 2007; 7(3):184–193.

Martin AM, Fishman R, Baxter L, Ford T. Practitioners' attitudes towards the use of standardized diagnostic assessment in routine practice: a qualitative study in two child and adolescent mental health services. Clinical Child Psychology and Psychiatry. 2011; 16(3):407–420. [PubMed: 21482583]

Mazonson PD, Mathias SD, Fifer SK, Buesching DP, Malek P, Patrick DL. The mental health patient profile: does it change primary care physicians' practice patterns? The Journal of the American Board of Family Practice. 1996; 9(5):336–345. [PubMed: 8884672]

Moløkken KJ, Jørgensen M. A Review of Surveys on Software Effort Estimation International Symposium on Empirical Software Engineering (ISESE 2003). 2003 Sep.

Mortenson BP, Witt JC. The use of weekly performance feedback to increase teacher implementation of a pre-referral academic intervention. School Psychology Review. 1998; 27:613–627.

New Freedom Commission on Mental Health. Achieving the Promise: Transforming Mental Health Care in America. Rockville, MD: 2003. Final Report. DHHS Pub. No. SMA-03-3832

Noell, GH.; Gansle, KA. The use of performance feedback to improve intervention implementation in schools. In: Sanetti, LM Hagermoser; Kratochwill, TR., editors. Treatment integrity: A foundation for evidence-based practice in applied psychology. Washington, DC: American Psychological Association; 2014. p. 161-184.

Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC. A compilation of strategies for implementing clinical innovations in health and mental health. Medical Care Research and Review. 2012; 69(2):123–157. [PubMed: 22203646]

Riemer, M.; Bickman, L. Using program theory to link social psychology and program evaluation. In: Mark, MM.; Donaldson, SI.; Campbell, B., editors. Social Psychology and Program/Policy Evaluation. New York: Guilford; 2011.

Riemer M, Rosof-Williams J, Bickman L. Theories related to changing clinician practice. Child and Adolescent Psychiatric Clinics of North America. 2005; 14:241–254. [PubMed: 15694784]

Robinson MB, Thompson E, Black NA. Evaluation of the effectiveness of guidelines, audit and feedback: improving the use of intravenous thrombolysis in patients with suspected acute myocardial infarction. International Journal for Quality in Health Care. 1996; 8(3):211–222. [PubMed: 8885185]

Rokstad K, Straand J, Fugelli P. Can drug treatment be improved by feedback on prescribing profiles combined with therapeutic recommendations? A prospective, controlled trial in general practice. Journal of Clinical Epidemiology. 1995; 48:1061–1068. [PubMed: 7775993]

Rose DJ, Church RJ. Learning to teach: the acquisition and maintenance of teaching skills. Journal of Behavioral Education. 1998; 8(1):5–35.

Rotheram-Borus MJ, Swendeman D, Chorpita BF. Disruptive innovations for designing and diffusing evidence-based interventions. American Psychologist. 2012; 67(6):463–476. [PubMed: 22545596]

Sapyta J, Riemer M, Bickman L. Feedback to clinicians: Theory, research, and practice. Journal of Clinical Psychology. 2005; 61(2):145–153. [PubMed: 15609360]

Southam-Gerow MA, Prinstein MJ. Evidence base updates: The evolution of the evaluation of psychological treatments for children and adolescents. Journal of Clinical Child and Adolescent Psychology. 2014; 43(1):1–6. [PubMed: 24294989]

Tabenkin H, Steinmetz D, Eilat Z, Heman N, Dagan B, Epstein L. A peer review programme to audit the management of hypertensive patients in family practices in Israel. Family Practice. 1995; 12(3):309–312. [PubMed: 8536836]

Tuckman BW, Yates D. Evaluating the student feedback strategy for changing teacher style. The Journal of Educational Research. 1980; 74(2):74–77.

Useem J. Business school, disrupted. The New York Times. 2014 May 31.

Wager, KA.; Lee, FW.; Glaser, JP. Health Care Information Systems: A Practical Approach for Health Care Management—Second Edition. San-Francisco: Jossey-Bass; 2009.

Weisz JR, Ng MY, Bearman SK. Odd couple? Reenvisioning the relation between science and practice in the dissemination-implementation era. Clinical Psychological Science. 2014; 2(1):58–74.

Wolpert M, Curtis-Tyler K, Edbrooke-Childs J. A qualitative exploration of patient and clinician views on patient reported outcome measures in child mental health and diabetes services. Administration and Policy in Mental Health and Mental Health Services Research. 2014

Fig. 1.

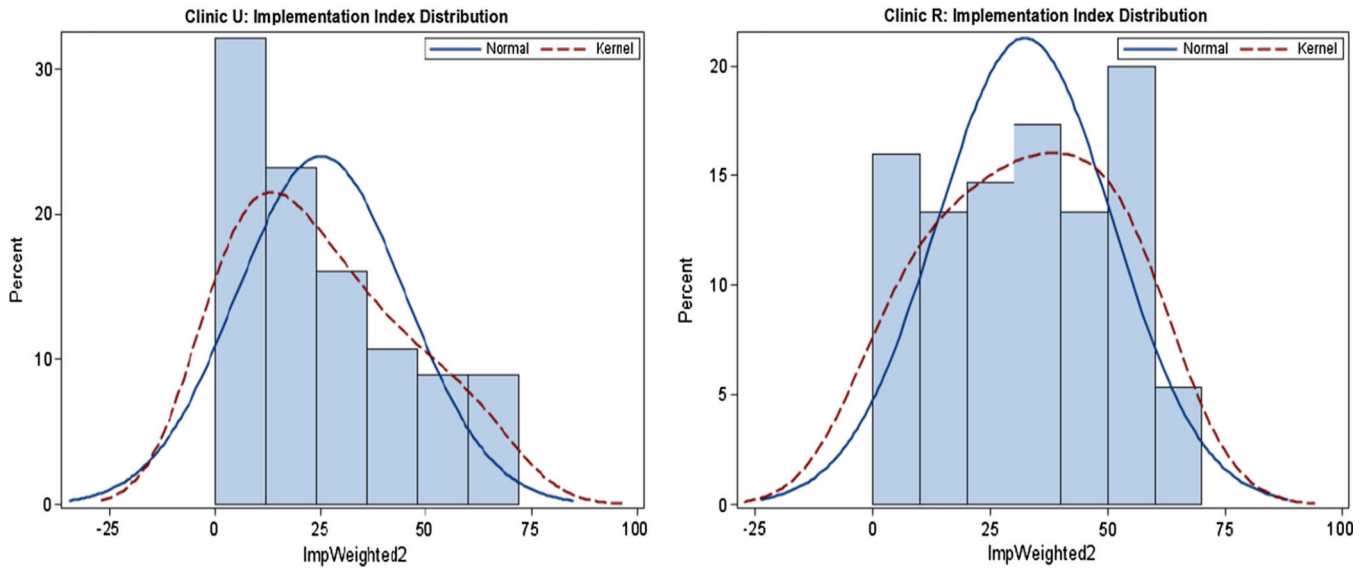Distribution of Youth's implementation index in Clinic U (*left panel*) and Clinic R (*right panel*)

Table 1

Youths and sessions by site and experimental condition

Site	Experimental Condition		Total
Clinic R	Feedback	Control	
N Youths	76	65	141
N Sessions	732	708	1440
Clinic U			
N Youths	56	60	116
N Sessions	662	596	1258

Table 2

Baseline characteristics of youths and caregivers by treatment Group within site

Characteristic	Clinic U							Clinic R						
	Control		Feedback		Test Statistic	p	Adj p*	Control		Feedback		Test Statistic	p	Adj p*
	N	%	N	%				N	%	N	%			
Youth's Characteristics														
Female	34	59	32	60	χ^2=0.04	0.85	1.00	31	50	30	43	χ^2=0.56	0.46	1.00
Ethnicity														
Hispanic	24	43	21	44	χ^2=0.01	0.93	1.00	1	2	6	9	χ^2=3.08	0.08	0.84
African American	28	52	26	54	χ^2=0.05	0.82	1.00	1	2	2	3	χ^2=0.23	0.63	1.00
White	2	4	1	2	χ^2=0.23	0.63	1.00	56	93	56	82	χ^2=3.51	0.06	0.78
Asian	1	2	-	-	χ^2=0.90	0.34	1.00	-	-	-	-			
Other Ethnic	23	43	20	42	χ^2=0.01	0.92	1.00	3	5	8	12	χ^2=1.86	0.17	1.00
Caregiver's Characteristics														
Female	50	86	50	94	χ^2=2.05	0.15	1.00	51	88	55	83	χ^2=0.53	0.47	1.00
Marital Status														
Single	3	7	1	3	χ^2=0.57	0.45	1.00	-	-	2	4	χ^2=2.00	0.16	0.99
Married/live as Married	23	50	14	40	χ^2=0.80	0.37	1.00	31	58	30	56	χ^2=0.09	0.76	1.00
Other	20	43	20	57	χ^2=1.48	0.22	1.00	22	42	22	41	χ^2=0.01	0.94	1.00
Relationship to Youth														
Birth Parent	50	88	35	70	χ^2=5.12	0.02	0.64	51	91	58	91	χ^2=0.01	0.93	1.00
Grandparent	-	-	8	16	χ^2=9.86	0.00	0.11	1	2	2	3	χ^2=0.22	0.64	1.00
Other Family	4	7	5	10	χ^2=0.31	0.58	1.00	4	7	4	6	χ^2=0.04	0.84	1.00
Foster Parent	3	5	-	-	χ^2=2.71	0.10	0.97	-	-	-	-			
Youth Age (M±SD)	14.3±1.94		14.0±2.18		z=−1.00	0.31	1.00	14.6±2.19		14.53±2.30		z=0.36	0.72	1.00
SFSS Intake Scores														
Youth SFSS (M±SD)	2.30±0.54		2.45±0.71		z=1.33	0.18	1.00	2.45±0.70		2.42±0.59		z=−0.28	0.78	1.00
Caregiver SFSS (M±SD)	2.58±0.68		2.56±0.77		z=0.07	0.94	1.00	2.42±0.71		2.47±0.74		z=0.27	0.79	1.00

Dash (−) indicates no youth;

*bootstrapped

Table 3

Characteristics of youth treatment episode by treatment group within site

Characteristic	Clinic U							Clinic R						
	Control		Feedback		Test Statistic	p	Adj. p*	Control		Feedback		Test Statistic	p	Adj. p*
	Mean	SD	Mean	SD				Mean	SD	Mean	SD			
Length in Tx (weeks)	17.23	13.49	19.82	17.65	z =0.35	0.72	1.00	17	13.05	16.76	14.1	z =-0.26	0.79	1.00
Session attendance (from billing data)														
Total sessions held	10.39	7.85	12.96	13.08	z =0.57	0.57	1.00	10.73	8.24	9.74	7.45	z =-0.20	0.84	1.00
Total sessions not held	3.08	2.84	3.56	3.79	z =0.54	0.59	1.00	2.16	2.08	1.69	2.12	z =2.00	0.05	1.00
% sessions held	0.79		0.79		z =0.07	0.94	1.00	0.83		0.87		z =-1.98	0.05	0.66
% Sessions not held	0.21		0.21		z =-0.08	0.93	1.00	0.17		0.13		z =1.96	0.05	0.68
Sessions with completed SFSS measures														
Count youths' SFSS	5.23	5.11	7.95	9.52	z =1.43	0.15	0.93	6.88	6.11	6.34	6.91	z =1.06	0.29	1.00
Count caregivers' SFSS	3.08	3.15	4.55	8.30	z =-0.12	0.90	1.00	4.34	5.26	3.74	3.75	z =0.24	0.81	1.00
Count therapists' SFSS	4.93	5.18	7.67	10.77	z =0.79	0.43	0.97	7.79	6.75	7.28	6.51	z =-0.16	0.88	1.00
Youth SFSS completion %	54.42		60.33		z =1.06	0.29	1.00	67.25		63.15		z =1.03	0.30	1.00
Caregivers SFSS completion %	38.33		35.39		z =-0.58	0.56	1.00	43.61		43.96		z =-0.04	0.97	1.00
Clinician SFSS completion %	41.91		44.57		z =0.43	0.67	1.00	67.32		70.23		z =-0.65	0.51	1.00
N therapists × youth	1.75	0.60	1.71	0.46	z =-0.16	0.87	1.00	1.99	1.05	1.73	0.85	z =1.33	0.18	0.90

*Bootstrapped

Table 4

Mode of data entry (%) by site and treatment group

| | Clinic U | | | | | Clinic R | | | | |
| | Control | | Feedback | | p value | Control | | Feedback | | p value |
	Manual	Electronic	Manual	Electronic		Manual	Electronic	Manual	Electronic	
Youth SFSS	61.58	38.42	59.60	40.40	0.79	14.63	85.37	17.10	82.90	0.58
Caregiver SFSS	75.55	24.45	81.36	18.64	0.38	41.34	58.66	38.20	61.80	0.57
Clinician SFSS	0.69	99.31	5.68	94.32	0.12	1.68	98.32	3.53	96.47	0.25

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

Table 5

ITT Results from HLM for Clinic U (Ns listed as Sessions, Youths, Clinicians)

	Youth SFSS (Ns: 756, 115, 11)			Caregiver SFSS (Ns: 438, 113, 11)			Clinician SFSS (Ns: 723, 94, 11)		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Fixed Effects									
Intercept	2.21	0.07	.000 *	2.53	0.08	.000 *	2.26	0.07	.000 *
Feedback	0.11	0.10	.274	0.06	0.12	.631	−0.00	0.10	.978
Slope (weeks after Tx starred)									
Intercept	−0.02	0.00	.000 *	−0.01	0.00	.049 **	−0.01	0.00	.053
Feedback * Slope	−0.00	0.00	.858	0.01	0.00	.129	0.00	0.00	.671
Random									
UN(1,1)	0.227	0.04	.000 *	0.309	0.06	.000 *	0.17	0.03	.000 *
Residual	0.187	0.01	.000 *	0.173	0.01	.000 *	0.14	0.01	.000 *

** *p* < .05;

* *p* < .0001

Table 6

ITT results from HLM for clinic R (Ns listed as sessions, youths, clinicians)

	Youth SFSS (Ns: 930, 141, 13)			Caregiver SFSS (Ns: 568, 141, 13)			Clinician SFSS (Ns: 1061, 138, 13)		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Fixed Effects									
Intercept	2.23	0.07	.000 *	2.29	0.08	.000 *	2.18	0.04	.000 *
Feedback	−0.05	0.09	.634	0.06	0.11	.592	0.10	0.06	.094
Slope (weeks after Tx started)									
Intercept	−0.02	0.00	.000 *	−0.01	0.00	.001 **	−0.00	0.00	.085
Feedback * Slope	0.00	0.00	.249	−0.01	0.01	.065	−0.01	0.00	.045
Random									
UN(1,1)	0.27	0.04	.000 *	0.34	0.05	.000 *	0.09	0.01	.000 *
Residual	0.17	0.01	.000 *	0.18	0.01	.000 *	0.13	0.01	.000 *

*
$p < .0001$;

**
$p < .05$

Table 7

Description, Explanation and Interpretation of Implementation Index examples

Implementation Index	Examples of how Implementation Index is obtained[a]	Interpretation
100	Each respondent (n=3) completed a questionnaire at every session and the clinician viewed every resulting feedback report	Ideal implementation: this client received 100% of the intended intervention
50	Each respondent completed questionnaires at half of the clinical sessions and the clinician viewed every resulting feedback report. OR Each respondent completed a questionnaire at every session and the clinician only viewed half of the resulting feedback reports OR Respondents completed questionnaires at an average of 90% of all clinical sessions and the clinician viewed 56% of resulting feedback reports.	Partial implementation: this client received 50% of the intended intervention
0	No respondent completed questionnaires at any clinical session The clinician did not view any feedback reports when questionnaires were completed by any respondent at any session	Complete implementation failure: This client received none of the intended intervention (i.e., equivalent to the control group).

[a]This column is not exhaustive of examples for obtaining an implementation index of 50

Table 8

Feedback condition: average questionnaire completion, feedback viewing across respondents, and resulting implementation indices by clinic

	Clinic U	Clinic R	Clinic R - U diff
Questionnaire completion			
Youth SFSS	63	63	0
Caregiver SFSS	37	44	7
Clinician SFSS	47	70	23
Feedback viewing			
Youth SFSS	71	63	−8
Caregiver SFSS	61	59	−2
Clinician SFSS	45	67	22
Implementation index			
Youth SFSS	45	40	−5
Caregiver SFSS	23	26	3
Clinician SFSS	21	47	26
Mean implementation index [*]	27	34	7

[*]Average across three respondents: youth, caregivers and clinicians. Mean Clinic R - U differential p value = 0.037

Table 9

HLM results investigating dose-effect for feedback group in Clinic U (Ns listed as sessions, youths, clinicians)

Clinic U	Youth SFSS (Ns: 439, 56, 11)			Caregiver SFSS (Ns: 252, 53, 11)			Clinician SFSS (Ns: 425, 44, 11)		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Fixed Effects									
Intercept	2.34	0.13	<.0001	2.37	0.18	<.0001	2.19	0.12	<.0001
Implementation	−0.00	0.00	0.940	0.01	0.01	0.215	0.00	0.00	0.527
Slope (weeks after Tx started)									
Intercept (time)	−0.02	0.00	<.0001	0.00	0.01	0.672	−0.01	0.00	0.006
Implementation Index*SloPe	−0.00	0.00	0.896	−0.00	0.00	0.444	0.00	0.00	0.059
Random									
UN(1,1)	0.29	0.07	<.0001	0.44	0.11	<.0001	0.16	0.04	<.0001
Residual	0.18	0.01	<.0001	0.15	0.02	<.0001	0.14	0.01	<.0001

Table 10

HLM results investigating Dose-Effect for Feedback Group in Clinic R (N's listed as sessions, youths, clinicians)

	Youth SFSS (Ns: 474, 75, 13)			Caregiver SFSS (Ns: 279, 75, 13)			Clinician SFSS (Ns: 544, 75, 13)		
	Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Fixed Effects									
Intercept	2.19	0.14	<.0001	2.37	0.17	<.0001	2.36	0.08	<.0001
Implement	−0.00	0.00	0.990	−0.00	0.01	0.919	−0.002	0.00	0.263
Slope (Weeks after Tx started)									
Intercept	0.018	0.01	0.007	−0.013	0.02	0.401	0.004	0.01	0.478
Implementation Index*Slope	−0.001	0.00	<.0001	−0.0002	0.00	0.683	−0.0003	0.00	0.021
Random									
UN(1,1)	0.30	0.06	<.0001	0.34	0.08	<.0001	0.08	0.02	<.0001
Residual	0.17	0.01	<.0001	0.20	0.02	<.0001	0.15	0.01	<.0001

Table 11

Effect sizes for dose-effect of feedback in Clinic R

Weeks[a]	Youth SFSS				Clinician SFSS			
	Implementation Index				Implementation Index			
	0	25	50	75	0	25	50	75
8	−0.25	0.02	0.29	0.57	−0.13	0.13	0.39	0.64
10	−0.31	0.03	0.37	0.71	−0.16	0.16	0.48	0.80
15	−0.47	0.04	0.55	1.06	−0.25	0.24	0.72	1.21

[a]weeks in treatment phase