Cultivating Evidence-Based Clinical Reasoning and Action in Youth Mental Health Care: The Reaching Families Multisite Randomized Trial

Bruce F. Chorpita[1], Kimberly D. Becker[2], Alayna L. Park[3], Davielle Lakind[4], Karen Guan[5],
Maya M. Boustani[6], Meredith R. Boyd[7], Wendy Chu[2], Eleanor G. Wu[2], and Kendra S. Knudsen[1]

[1] Department of Psychology, University of California, Los Angeles
[2] Department of Psychology, University of South Carolina
[3] Department of Psychology, University of Oregon
[4] Department of Clinical Psychology, Mercer University
[5] Pacific Clinics, Los Gatos, California, United States
[6] Department of Psychology, Loma Linda University
[7] Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine

Objective: Despite decades of policy emphasizing the role of evidence in guiding services, few studies have sought to improve the degree to which evidence is used in supervision and treatment. This study reports supervisor and therapist outcomes from the Reaching Families multisite cluster-randomized controlled trial, which tested the effects of a coordinated knowledge system (CKS) against practice guidelines (PG) on the use of evidence in supervision and treatment targeting low treatment engagement in publicly funded youth community mental health organizations located in two geographically distinct, underresourced communities where service inequities are common. *Method:* The sample included 121 mental health professionals (92.6% female; 81.0% Black, Indigenous, and people of color[1]) randomly assigned to a CKS or PG control condition. We recorded, transcribed, and coded 430 supervision and 208 treatment sessions involving 221 youth (M_{age} = 13.1 years, 46.2% female; 78.7% Black, Indigenous, and people of color) and/or their caregivers who reported engagement concerns during therapy. *Results:* CKS dyads showed uniformly greater use of evidence focused on specific client needs relative to dyads in the PG condition, with large effect sizes and no differences in the effect of condition across the sites. Secondary analyses showed that tools in the CKS condition were perceived significantly more positively than those in the PG condition in terms of effort and effectiveness, and supervisory workload was the same across both conditions. *Conclusions:* In routine clinical care delivered within highly representative community settings, a strategically designed knowledge resource can improve evidence-based reasoning and action and be perceived as easy to use and useful without negatively impacting workload.

What is the public health significance of this article?
Mental health care involves many complex decisions that lack the tools to inform them. This study found that, relative to a practice guidelines control condition, a resource system could improve clinical reasoning and use of relevant evidence in decision making with relatively light training and effort.

Keywords: evidence-based practice, clinical reasoning, mental health services, supervision, engagement

Psychology has long recognized the reciprocal benefits of science and practice (e.g., Thorne, 1947; Witmer, 1907/1996, reprinted 1996) and, nearly 2 decades ago, declared a strong commitment to evidence-based practice in psychology (EBPP; APA Presidential Task Force on Evidence-Based Practice, 2006). EBPP has been described as a methodical approach to clinical reasoning and action, involving the integration of the best available evidence (e.g., research, theory, case-based data; Rousseau & Gunia, 2016) with clinical expertise and

patient values (APA Presidential Task Force on Evidence-Based Practice, 2006; Sackett et al., 1996, 2000). EBPP emerged in the

---

[1] BIPOC refers to Black, Indigenous, and people of color. Following the guidance of the District of Columbia's Office of Human Rights and the Mayor's Office of Racial Equity (2022), we acknowledge that using a single term to represent diverse racial and ethnic backgrounds is inherently limiting. Therefore, we have restricted our use of this term to the abstract of this article.

Jasper Smits served as action editor.
Bruce F. Chorpita https://orcid.org/0000-0001-7454-464X
Kimberly D. Becker https://orcid.org/0000-0003-2381-8588
Alayna L. Park https://orcid.org/0000-0001-8869-4438
Davielle Lakind https://orcid.org/0000-0001-7558-9989
Karen Guan https://orcid.org/0000-0002-7312-8078

Maya M. Boustani https://orcid.org/0000-0002-7739-7857
Meredith R. Boyd https://orcid.org/0000-0001-7736-3430
Wendy Chu https://orcid.org/0000-0001-6596-5987
Eleanor G. Wu https://orcid.org/0000-0003-2227-3176
Kendra S. Knudsen https://orcid.org/0000-0002-3516-8917

medical field in the 1980s (as evidence-based medicine; Djulbegovic & Guyatt, 2017; Sackett, 1981, 1989), spurred by the recognition that a lack of practice standards led to inconsistent and questionable treatment quality (Barends & Briner, 2014; Rousseau & Gunia, 2016).

In roughly this same historical period, there has been an adjacent conceptualization prioritizing the notion of evidence-based treatment (EBT), which has often been described as high-fidelity application of manualized interventions that have strong empirical support (e.g., Chambless & Hollon, 1998). In children's mental health alone, the EBT paradigm has yielded hundreds of effective interventions for a wide variety of clinical concerns (Chorpita et al., 2011; Weisz et al., 2023), and the EBT conceptualization, rather than EBPP, has been the dominant paradigm in defining practice within research and implementation policy contexts for over 30 years.

Despite this dominance of EBT in practice policy, the implementation of EBTs in community settings has a long history of challenges, which include a documented need for real-time adjustments and adaptations (e.g., Guan et al., 2019; Henggeler & Schoenwald, 2002; Schoenwald et al., 2011). This specific barrier has led to calls for more diverse characterizations of psychosocial interventions in clinical trials and implementation research that are not limited to highly structured manualized interventions (e.g., Wiltsey-Stirman & Comer, 2018). Despite these calls, however, operationalizing how mental health professionals (MHPs) reason using evidence when delivering care has garnered relatively little attention compared with the measurement of EBT fidelity (see Real & Poole, 2005; Rotheram-Borus et al., 2012). That said, some leading scholars and EBT developers have offered characterizations of the "thinking behind the action" in the use of EBTs (e.g., Chambers & Norton, 2016; Kendall & Chu, 2000), continuing to emphasize the importance of strategic reasoning in intervention, above and beyond the tactical implementation of structured procedures.

With that in mind, we sought to (a) operationalize clinical reasoning and service activities that incorporate case-based and research-derived evidence and (b) examine whether we could increase the rate of evidence-based reasoning and action among MHPs seeking to improve treatment engagement in a community-based children's mental health context. Of relevance is a growing literature on how and when research evidence is used in practice, referred to as "use of research evidence" (URE; e.g., Crowley et al., 2021; Rickinson et al., 2022; Supplee et al., 2023). To operationalize these activities in clinical care, we drew from knowledge-to-action models (e.g., Deming, 1993; Graham et al., 2006), which specify a course of reasoning that includes detecting and prioritizing problems, considering and selecting responses in the context of their research support and fit to the case, implementing those solutions, and then reviewing both their integrity and impact. For example, Graham et al. (2006) outlined a knowledge-to-action cycle that included problem identification, selection of interventions, adapting or preparing for barriers, implementation of the interventions, and monitoring of implementation quality and expected outcomes.

In the URE literature, such decision and action flows are often considered in a social context, with an emphasis on management processes (e.g., supervision) to promote integrity and accountability;

Correspondence concerning this article should be addressed to Bruce F. Chorpita, Department of Psychology, University of California, Los Angeles, Box 951563, Los Angeles, CA 90095, United States. Email: chorpita@ucla.edu

thus, we designed our investigation to include the context of clinical supervision. Finally, this operationalization of evidence-based reasoning and action acknowledges the role of knowledge resources (or "knowledge tools"; Graham et al., 2006, p. 19), which deliver consolidated evidence summaries relevant to a specific scenario or activity, such as assessment, planning, or treatment delivery. Hence, we designed a set of task-specific tools that would serve to organize the evidence base for mental health providers, so that we could examine the effects of these tools on their evidence-based reasoning and action.

For this trial, we chose a variety of intersecting contexts conducive to testing our operationalization of evidence-informed clinical reasoning and service activities. In particular, we sought a context in which clinical complexity would be high and the availability of evidence to drive decisions and actions would be comparatively low, thus elevating clinical uncertainty and the need for on-the-fly reasoning that can often be a part of community care (Garland et al., 2010; Guan et al., 2019; Wiltsey-Stirman et al., 2017). We chose a context of school-based mental health services, which is highly representative of the complications that arise in routine care (Southam-Gerow et al., 2008) and is the most common context for mental health services for youth (Duong et al., 2021). We chose sites that were geographically and culturally diverse, with vastly different histories of public investment in their workforce capacity, and we partnered with local MHPs who represented the expertise and values of their communities. Finally, we chose to focus on evidence-based reasoning and action as it pertains to low treatment engagement, given the pervasiveness of treatment engagement challenges in mental health services (e.g., Pellerin et al., 2010; Saloner et al., 2014), the relative paucity of clinical trial research on engagement (Becker & Chorpita, 2023; Becker et al., 2018), and findings that MHPs generally have limited knowledge of the research evidence on treatment engagement, regardless of their training history or experience using EBTs for youth mental health concerns (Reeder et al., 2024), and are thus generally underprepared to address engagement challenges effectively.

Our primary aims were to test whether we could intervene with supervisors and therapists to engage in a process of reasoning and action that spanned multiple supervision and treatment events. This process included evidence-informed reflection on engagement concerns, selection of a specific immediate target focus, consideration of relevant engagement procedures that had research support, selection of and preparation to deliver procedures judged to be a good fit for the youth and family, the application of those procedures in treatment, and a review of their integrity and impact. Specifically, our first primary aim was to detect whether this reasoning process could be improved in supervision through the use of a coordinated knowledge system (CKS), and our second primary aim was to see whether this system would produce concurrent improvement in the application of relevant evidence-based procedures in treatment. We hypothesized that the CKS would have a significant positive effect, which would be comparable across service systems with considerably different cultural, geographic, and service infrastructure characteristics. Finally, because we were also concerned with the sustainability of improved use of evidence in reasoning and clinical care in this context, we pursued secondary aims to examine MHPs' perceived level of effort and utility, using self-report measures as well as overall work volume. We expected

the CKS would demonstrate significantly higher utility and lower perceived effort, due to its design aims to organize relevant evidence and expedite decision making and planning. For work volume, we expected comparable findings across conditions (i.e., there would not be significantly greater demand for MHPs in the CKS condition).

Importantly, we did not investigate youth and caregiver engagement outcomes in this study. Rather, this study was designed explicitly to test the capacity of a knowledge system to produce MHP reasoning and action that incorporated case-based and research-based evidence, across a focused set of clinical events. A trial to produce youth or caregiver engagement outcomes could require a dose of engagement practices lasting more than one or two sessions, and the integrity and impact of those practices could benefit from ongoing consultation and measurement, which we did not provide. As designed, the present study provides a strong and innovative test of whether it is possible to cultivate evidence-based reasoning and action in a highly challenging context demanding real-time decisions about emergent treatment engagement challenges.

Method

All procedures for this study, referred to as the Reaching Families trial, were reviewed and approved by the Institutional Review Boards of the University of California, Los Angeles; the University of South Carolina; as well as the participating service organizations that requested independent reviews. Below we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Reaching Families Study Context and Participants

Participating mental health service systems were located in Los Angeles, California (CA) and in the Pee Dee and Santee Wateree areas of South Carolina (SC). School mental health (SMH) services in CA were delivered by professionals employed by the Los Angeles Unified School District SMH Clinic and Wellness Center program. SMH services in SC were delivered by contracted professionals employed by the SC Department of Mental Health. Both programs offer a variety of multitiered supports (e.g., mental health promotion, prevention, and intervention services) in an array of formats (e.g., classroom presentations, individual and group therapies) and are located in underresourced communities where service inequities are common.

Services for youth behavioral and emotional problems were delivered using the locally prioritized EBTs in each setting. Due to a 12-year county-wide investment in EBTs and workforce development in CA (e.g., Regan et al., 2017), training in EBTs was more prevalent there than in SC, where such investments were minimal. In general, weekly individual appointments with youth were standard, and both sites offered flexible scheduling (e.g., every other week) based on youth needs and preferences. Overall, these therapists and their services were comparable to a typical outpatient master's level workforce, although access to youth clients was generally higher in these school-based contexts than in other community-based outpatient settings. In the services delivered in CA and SC, caregivers were invited to participate in treatment. Our review of service records across a 2-year interval at both sites showed that caregivers attended approximately one third of treatment sessions. In addition, in

approximately half of treatment cases, caregivers were involved in at least one of the first four sessions.

MHP Participants

From each site, we recruited MHPs working within an SMH service context. MHPs were eligible if they had the role of mental health therapist or supervisor and the ability to read and speak English fluently. There were no exclusion criteria for MHPs. They were invited to participate in the study through initial conversations with administrators at both sites and then through in-person events local to each service system in which study procedures were fully explained and informed consent was obtained. Figure 1 shows a Consolidated Standards of Reporting Trials flow (Campbell et al., 2004) representing the enrollment of MHP participants into the study. Overall, consent rates were extremely high. Rates of MHP attrition were low, with fewer than 10% of MHPs leaving the study before having a qualifying youth case. The sample size for MHPs

Figure 1

CONSORT Flow for Mental Health Professionals
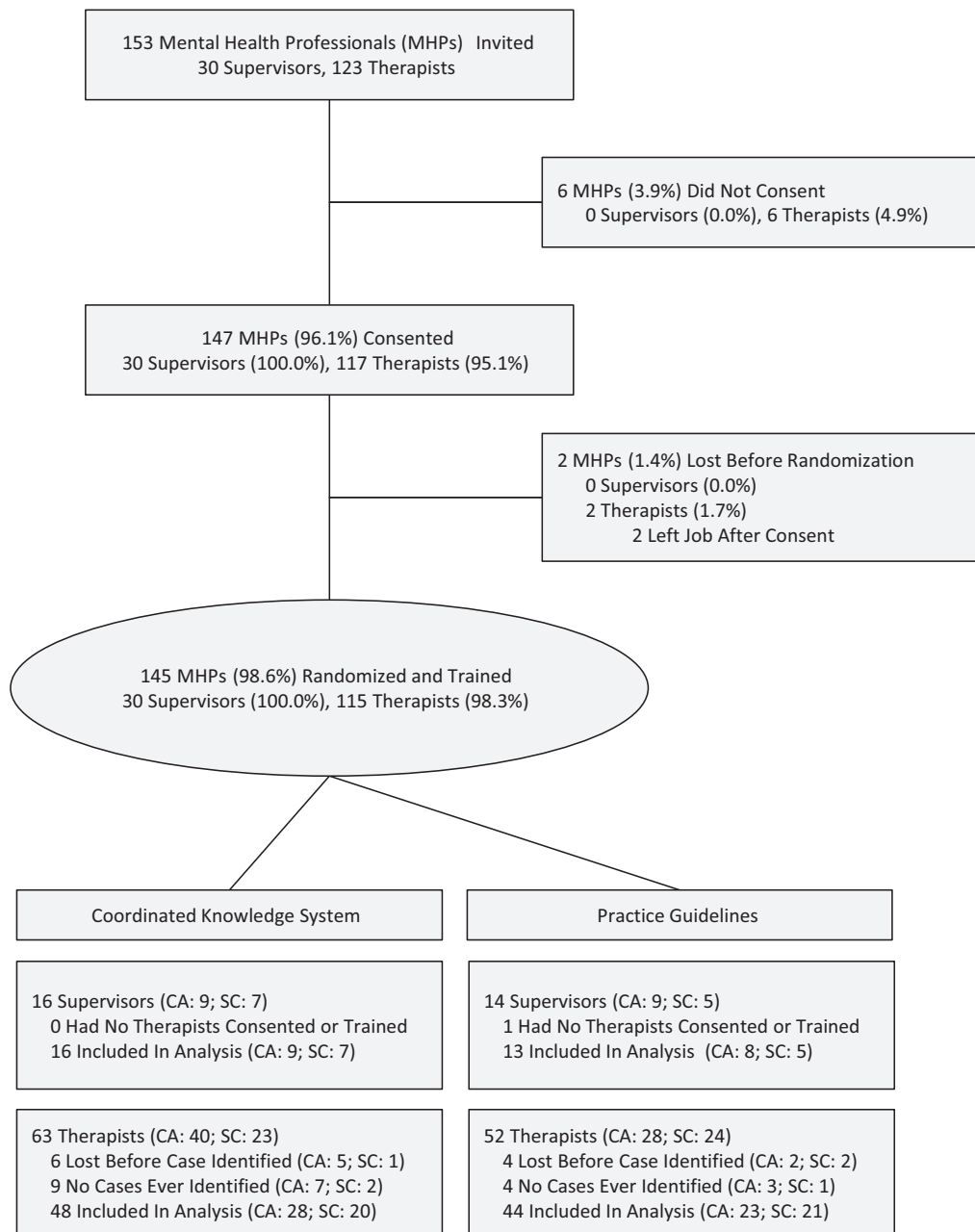
Note. CONSORT = Consolidated Standards of Reporting Trials; CA = California; SC = South Carolina.

was based on initial enrollment targets to achieve sufficient statistical power and did not include those who dropped out of the study before screening a positive youth case or who had no youth cases screened positive for the duration of the study.

Supervisors. Table 1 displays the background characteristics of 29 supervisors from CA ($n = 17$; 58.62%) and SC ($n = 12$; 41.38%) who contributed youth cases to the study. The supervisory workforce predominantly consisted of women in their middle years from ethnic–racial minoritized groups. Most possessed master's degrees, described their theoretical orientation as either cognitive behavioral or eclectic, and had dedicated almost as many years to their organization as they had spent in the field of mental health services, showcasing an enduring commitment to their workplace. Tests for differences in variables in Table 1 showed no significant differences across conditions; however, some site differences emerged using Fisher's exact test. Specifically, there was an overall site effect for race, $\chi^2(3) = 11.33$, $p < .01$, with significantly more African American/Black supervisors at the SC site, $\chi^2(1) = 10.08$, $p < .01$, and an overall site effect for educational background, $\chi^2(3) = 25.12$, $p < .01$, with significantly more CA supervisors having a Master of Social Work, $\chi^2(1) = 21.34$, $p < .001$, and significantly more SC supervisors having an MA or MS degree in a field other than social work, $\chi^2(1) = 21.62$, $p < .001$. Reflecting differences in policy and workforce capacity, CA supervisors were significantly more likely to be licensed, $\chi^2(1) = 25.11$, $p < .001$, with only a single supervisor in SC reporting holding a state license in their profession. There were no site differences for theoretical orientation.

Therapists. Table 2 displays the background characteristics of 92 therapists from CA ($n = 51$; 55.43%) and SC ($n = 41$; 44.57%) with participating cases in the study. In general, therapist participants predominantly consisted of women in their middle years from ethnic–racial minoritized groups. There were no significant differences observed across study conditions on any background variables. However, several distinctions based on site were evident. Notably, therapists from the SC site were significantly older, $F(1, 82) = 13.76$, $p < .001$, but had similar years since obtaining their degree and experience in their current organizations, suggesting that professional credentials were obtained later in life than in CA. The average caseload was nearly double in SC, $F(1, 89) = 24.83$, $p < .001$, an indicator of the lower workforce availability at that site. There was also a site effect for race/ethnicity, $\chi^2(3) = 65.11$, $p < .001$, with significantly more Latine therapists in CA, $\chi^2(1) = 55.67$, $p < .001$, and significantly more African American/Black therapists in SC, $\chi^2(1) = 53.11$, $p < .001$. Consistent with the different demands within their respective systems, therapists in CA were significantly more likely to speak Spanish, $\chi^2(1) = 33.53$, $p < .001$, with nearly 90% of therapists there reporting fluency or proficiency for conducting therapy. As with the supervisor sample, educational background differed by site, $\chi^2(3) = 64.24$, $p < .001$, with more Master of Social Work degrees in CA, $\chi^2(1) = 53.11$, $p < .001$, and more MA or MS degrees in fields other than social work in SC, $\chi^2(1) = 64.01$, $p < .001$. Licensure was obtained by 41.3% of therapists, with rates significantly higher in CA than in SC, $\chi^2(1) = 8.73$, $p < .01$. No site differences were observed for theoretical orientation, which was most often described as cognitive behavioral.

Youth Participants

Figure 2 shows the flow of youth cases. Inclusion criteria required cases to have (a) at least one individual (i.e., youth,

Table 1

Supervisor Background by Site

Characteristic	CA	SC	Total
Mean age (*SD*)	45.99 (6.64)	46.46 (13.50)	46.16 (9.50)
Mean years since degree (*SD*)	16.38 (4.34)	15.50 (10.70)	16.06 (7.17)
Mean years in organization (*SD*)	14.24 (4.71)	11.50 (9.35)	13.22 (6.76)
Gender *n* (%)			
Female	17 (100.00)	11 (91.67)	28 (96.55)
Male	0 (0.00)	1 (8.33)	1 (3.45)
Race *n* (%)*			
African American/Black*	4 (23.53)	10 (83.33)	14 (48.28)
Asian American	2 (11.76)	0 (0.00)	2 (6.90)
Latine	6 (35.29)	0 (0.00)	6 (20.69)
White	5 (29.41)	2 (16.67)	7 (24.14)
Speaks Spanish *n* (%)	5 (29.41)	0 (0.00)	5 (17.24)
Education *n* (%)*			
MSW**	16 (94.12)	1 (8.33)	17 (58.62)
MA/MS**	0 (0.00)	10 (83.33)	10 (34.48)
MEd	0 (0.00)	1 (8.33)	1 (3.45)
PhD	1 (5.88)	0 (0.00)	1 (3.45)
State licensed *n* (%)**	17 (100.00)	1 (8.33)	18 (66.67)
Orientation *n* (%)			
Cognitive behavioral	9 (52.94)	9 (75.00)	18 (62.07)
Eclectic	4 (23.53)	1 (8.33)	5 (17.24)
Humanistic/client Centered	2 (11.76)	0 (0.00)	2 (6.90)
Family systems	1 (5.88)	0 (0.00)	1 (3.45)
Psychodynamic	1 (5.88)	0 (0.00)	1 (3.45)
Not reported	0 (0.00)	2 (16.67)	2 (6.90)

Note. CA = California; SC = South Carolina.
* $p < .01$. ** $p < .001$.

Table 2
Therapist Background by Site

Characteristic	CA	SC	Total
Mean age in years (*SD*)	35.09 (7.86)	42.82 (10.80)	38.36 (9.93)
Mean years since degree (*SD*)	6.46 (5.82)	6.35 (5.36)	6.41 (5.59)
Years in organization (*SD*)	4.67 (5.40)	4.37 (3.86)	4.53 (4.76)
Caseload (*SD*)**	19.80 (15.40)	35.71 (16.83)	26.89 (17.83)
Gender *n* (%)			
Female	46 (90.20)	38 (92.68)	84 (91.30)
Male	5 (9.80)	3 (7.32)	8 (8.70)
Race *n* (%)**			
African American/Black**	3 (5.88)	33 (80.49)	36 (39.13)
Asian American	3 (5.88)	0 (0.00)	3 (3.26)
Latine**	41 (80.39)	1 (2.44)	42 (45.65)
White	5 (9.80)	7 (17.07)	12 (13.04)
Speaks Spanish *n* (%)**	45 (88.24)	12 (29.27)	57 (61.96)
Education *n* (%)**			
MSW**	48 (94.12)	8 (19.51)	56 (60.87)
MA/MS**	0 (0.00)	33 (80.49)	33 (35.87)
BA/BS	2 (3.92)	0 (0.00)	2 (2.17)
PhD	1 (1.96)	0 (0.00)	1 (1.09)
State licensed *n* (%)*	28 (54.90)	10 (24.39)	38 (41.30)
Orientation *n* (%)			
Cognitive behavioral	32 (62.75)	27 (65.85)	59 (64.13)
Eclectic	13 (25.49)	5 (12.20)	18 (19.57)
Humanistic/client centered	1 (1.96)	5 (12.20)	6 (6.52)
Family systems	4 (7.84)	1 (2.44)	5 (5.43)
Psychodynamic	1 (1.96)	0 (0.00)	1 (1.09)
Other	0 (0.00)	3 (7.32)	3 (3.26)

Note. CA = California; SC = South Carolina.
* $p < .01$. ** $p < .001$.

caregiver) able to complete the My Thoughts about Therapy (MTT; described further below) in English or Spanish as part of the initial screening for treatment engagement concerns and (b) at least one initial screening with a valid MTT scale score below 14 (see Table 3 for scores). Cases were ineligible if they had already participated in the study or were administered the MTT in a manner or at a time inconsistent with study procedures (e.g., as therapy was terminating). Cases were excluded if therapists already had multiple participants enrolled in the study (to manage overall demand) or if therapists were unavailable (e.g., due to professional leave or illness). Cases were invited to participate by therapists, who also administered informed consent/assent procedures. Of those invited, roughly two thirds agreed to participate, and baseline screening scores did not differ for caregiver or youth report for cases who consented versus refused to participate. Cases were analyzed if they had at least one codable supervision session, which represented 95.67% of cases. Cases without any codable treatment event (5.88% of cases in the analysis) were included using the logic of an intent-to-treat analysis (e.g., Gupta, 2011), such that therapists were rated as not performing any of the measured treatment activities for that case.

Table 4 displays the background characteristics of the 221 youth participants included in the analyses. Students did not differ significantly across study conditions on any of the variables reported but did differ significantly across sites on several of the variables. Students ranged in age from 5 to 21 (CA: 5–21; SC: 6–19) and ranged in grades from preschool to 12th grade at both sites. The median grade was eighth grade in CA and fifth grade in SC. Students in CA were slightly, but significantly, older, $F(1, 217) = 9.54$, $p < .01$. Gender

obtained from electronic health records were male and female in similarly equal proportions at both sites (only binary gender status was represented in the records provided). Race/ethnicity differed by site, $\chi^2(3) = 153.50$, $p < .001$, with primarily Latine students in CA, $\chi^2(1) = 147.20$, $p < .001$, and African American/Black students in SC, $\chi^2(1) = 57.52$, $p < .001$. SC also had a significantly higher proportion of White students, $\chi^2(1) = 39.52$, $p < .001$.

Youths' primary clinical focus was obtained from electronic health records. In general, a wide variety of problems were represented such that this sample appeared to generalize well to other school-based mental health contexts. The distribution of primary problems differed by site, $\chi^2(5) = 75.61$, $p < .001$, with CA having significantly higher rates of mood-related problems, $\chi^2(1) = 21.91$, $p < .001$, and trauma-related problems, $\chi^2(1) = 13.57$, $p < .001$, and SC having higher rates of attention-related problems, $\chi^2(1) = 147.20$, $p < .001$.

Because our study team had no direct contact with youth and families, we did not obtain detailed background information on caregivers (e.g., age, marital status). However, of the 221 cases available for analysis, 184 had at least one caregiver report on the MTT–Caregiver Version in the initial screening, of which 122 were obtained from mothers (66.30%), 25 from fathers (13.59%), 13 from grandparents (7.07%), and 24 (13.04%) from other sources (e.g., stepparent, foster parent, legal guardian). Caregivers participated in 76 (36.54%) of the 208 treatment sessions analyzed for this study.

Study Design

We used a cluster-randomized design, with supervisors assigned to one of two study conditions using blocked randomization.

Figure 2
CONSORT Flow for Youth and Caregivers

```
┌─────────────────────────────┐
│    1,352 Cases Screened     │
└─────────────────────────────┘
                │
                │          ┌──────────────────────────────────────────┐
                ├─────────▶│ 641 (47.4%) Did Not Screen Positive      │
                │          └──────────────────────────────────────────┘
┌─────────────────────────────┐
│     711 (52.6%) Cases       │
│     Potentially Eligible    │
└─────────────────────────────┘
                │          ┌──────────────────────────────────────────┐
                │          │ 133 (18.7%) Technically Ineligible       │
                │          │    63 (47.4%) School Calendar            │
                ├─────────▶│    20 (15.0%) Timing Error               │
                │          │    16 (12.0%) Administration Error       │
                │          │    16 (12.0%) Team Communication         │
                │          │    11 (8.3%) Missing Data                │
                │          │     7  (5.3%) Other (e.g., already       │
                │          │               completed study)          │
┌─────────────────────────────┐ └────────────────────────────────────┘
│    578 (81.3%) Cases        │
│    Eligible                 │
└─────────────────────────────┘
                │          ┌──────────────────────────────────────────┐
                │          │ 93 (16.2%) Excluded                      │
                ├─────────▶│    78 (83.9%) Therapist or Supervisor    │
                │          │               Capacity                   │
                │          │    15 (16.1%) Therapist Unavailable      │
                │          └──────────────────────────────────────────┘
                │          ┌──────────────────────────────────────────┐
                ├─────────▶│ 120 (20.8%) Invitation Not Completed     │
                │          └──────────────────────────────────────────┘
┌─────────────────────────────┐
│   365 (63.1%) Cases Invited │
└─────────────────────────────┘
                │          ┌──────────────────────────────────────────┐
                ├─────────▶│ 134 (36.7%) Refused Consent              │
                │          └──────────────────────────────────────────┘
```



```
        ╭──────────────────────────────────╮
        │  231 (62.9%) Youth Cases         │
        │  Allocated                       │
        ╰──────────────────────────────────╯
             ╱                    ╲
┌───────────────────────────┐  ┌───────────────────────────┐
│ Coordinated Knowledge     │  │ Practice Guidelines       │
│ System                    │  │                           │
└───────────────────────────┘  └───────────────────────────┘

┌───────────────────────────────────┐  ┌───────────────────────────────────┐
│ 122 (52.8%) Youths allocated      │  │ 109 (47.2%) Youths allocated      │
│   (CA: 67, SC: 55)                │  │   (CA: 52, SC: 57)                │
│ 4 Did not receive intervention    │  │ 6 Did not receive intervention    │
│   (CA: 1, SC: 3)                  │  │   (CA: 0, SC: 6)                  │
│ 118 Received any intervention      │  │ 103 Received any intervention      │
│   (CA: 66, SC: 52)                │  │   (CA: 52, SC: 51)                │
└───────────────────────────────────┘  └───────────────────────────────────┘

┌───────────────────────────────────┐  ┌───────────────────────────────────┐
│ 118 Youths available for analysis │  │ 103 Youths available for analysis │
│   341 Supervision Events          │  │   290 Supervision Events          │
│     118 Supervision 1             │  │     103 Supervision 1             │
│     111 Supervision 2             │  │      98 Supervision 2             │
│   231 Treatment Events            │  │   196 Treatment Events            │
│     109 Treatment 1               │  │      99 Treatment 1               │
└───────────────────────────────────┘  └───────────────────────────────────┘
```

Note. CONSORT = Consolidated Standards of Reporting Trials; CA = California; SC = South Carolina.

Therapists and their respective cases were thus automatically allocated to the condition of the supervisor. Our rationale was to protect against diffusion of the independent variable (Kazdin, 2023), given the possibility that supervisors would potentially use procedures from their preferred condition if randomization was at the level of the therapist and likewise for therapists if randomization occurred at the level of cases. We used a random number generator to produce a sequence of numbers representing a condition, with

Table 3
Baseline Scores by Informant on the MTT Engagement Scales

Measure	*M*	*SD*	% Positive
MTT youth (*n* = 203)			
Relationship	15.10	3.07	22.00
Expectancy	15.07	3.19	29.35
Attendance	13.01	2.64	69.54
Clarity	15.10	3.14	22.61
Homework	14.52	3.05	32.49
MTT caregiver (*n* = 184)			
Relationship	16.42	4.08	11.96
Expectancy	16.17	3.90	13.66
Attendance	14.42	3.26	38.67
Clarity	15.77	3.67	11.60
Homework	15.55	3.42	12.92

Note. A positive score was defined as having a scale value below 14. MTT = My Thoughts About Therapy.

block size as the entire cohort of participating supervisors within each service organization and an allocation ratio of 1:1. These numbers were assigned to supervisors, with allocation concealed through the use of ID numbers. All cases for participating therapists were screened at approximately the fourth week of therapy, and no tangible incentives for study participation were offered to youth, families, therapists, or supervisors.

Coordinated Knowledge System

The experimental condition used a CKS (see Chorpita & Daleiden, 2018), which in this context was focused on problems concerning treatment engagement (Becker et al., 2019). A CKS organizes one or more knowledge resources (see Graham et al., 2006) for capturing and/or communicating case-based and/or research-based evidence (e.g., questionnaires, dashboard reports, evidence summaries, clinical protocols). A CKS also leverages case-level evidence (e.g., questionnaire-derived scores, normative benchmarks) reflecting the perspective of relevant informants (e.g., youth, caregiver, provider).

These systems can include dynamic knowledge resources, such that summaries of relevant research evidence update as the literature evolves (see Chorpita & Daleiden, 2014). The coordinating aspects of the system bias users toward a sequence or flow of activities, which can include conditionality (e.g., if X arises, then consider Y) and operational definitions of those triggering conditions, while managing dependencies among users, resources, and activities (cf. Malone & Crowston, 1994).

With a focus on youth and caregiver treatment engagement, the CKS in this study was called the Reaching Families Engagement System (RFES; described below). MHPs in the CKS condition attended a 2-day training event, with a half day covering study procedures (e.g., eligibility notification, consent procedures, instructions for recording) and a day and a half on the use of the RFES. RFES training focused on (a) how to use specific measures to detect and prioritize engagement concerns, based on a multidimensional model of treatment engagement (Becker et al., 2018), (b) a review of written guides summarizing 11 discrete procedures with empirical support for engagement, and (c) rehearsal with a one-page worksheet to guide the process of supervision and to inform the selection of procedures with empirical support for various engagement problems. No follow-up clinical consultation was offered to therapists or supervisors.

We developed the RFES as a set of tools and guides to support planning and monitoring in supervision, as well as the application of specific engagement-focused procedures in treatment sessions. Supervisory activities targeted included recognizing engagement issues, choosing appropriate interventions for the specific problem at hand, preparing for intervention implementation, carrying out the intervention, and assessing its impact on engagement (see Becker et al., 2019). The RFES included (a) a graphic shared with MHPs indicating focal risks for engagement (e.g., *relationship* problem), (b) a worksheet with cues to prompt MHP collaboration on therapist perceptions of youth/caregiver engagement and on strategies supported by research evidence that could be relevant, and (c) a collection of 11 two-page guides with brief instructions on how to deliver engagement techniques. The creation of the RFES was based on the core process management models and practice guide

Table 4
Youth Participant Characteristics by Site

Characteristic	CA	SC	Total
Mean age (*SD*)*	13.46 (3.52)	12.02 (3.17)	13.08 (3.43)
Gender *n* (%)			
Female	56 (47.46)	46 (44.66)	102 (46.15)
Male	62 (52.54)	57 (55.34)	119 (53.85)
Race *n* (%)**			
African American/Black**	15 (12.71)	63 (61.76)	78 (35.45)
Asian American	2 (1.69)	0 (0.00)	2 (0.91)
Latine**	96 (81.36)	0 (0.00)	96 (43.64)
White**	5 (4.24)	39 (38.24)	44 (20.00)
Primary problem *n* (%)**			
Mood related**	45 (38.46)	11 (10.68)	56 (25.45)
Attention related**	1 (0.85)	43 (41.75)	44 (20.00)
Disruptive behavior related	22 (18.80)	20 (19.42)	42 (19.09)
Anxiety related	18 (15.38)	14 (13.59)	32 (14.55)
Trauma related**	22 (18.80)	3 (2.91)	25 (11.36)
Other related	9 (7.69)	12 (11.65)	21 (9.55)

Note. CA = California; SC = South Carolina.
* *p* < .01.   ** *p* < .001.

resources that are part of the Managing and Adapting Practice system (Chorpita & Daleiden, 2014); the RFES was thus meant to promote an intentional, evidence-informed, and self-correcting process of clinical reasoning and action. Training in the use of the RFES involved limited rehearsal of therapy activities and instead focused mainly on supervisor–therapist interaction.

Practice Guidelines

To provide a strong experimental test, we chose not to use a routine care control condition, instead building in two specific enhancements to service as usual. First, we alerted providers in the practice guidelines (PG) condition about each case in which either the youth or caregiver had reported an engagement concern. We did this because prior research has demonstrated that MHPs frequently underdetect engagement concerns (e.g., Becker et al., 2021; Westmacott & Hunsley, 2017), and thus, we would expect to see minimal engagement-relevant clinical reasoning for cases about which concerns were unknown to the treatment team. We, therefore, informed MHPs in this condition of each case that had reported an engagement concern, so that they could act at their discretion. MHPs in this condition did not have access to scores from the MTT questionnaires.

Second, we provided a set of written PG to all MHPs in this condition, which detailed the same 11 practices that were included in the RFES for MHPs in the CKS condition. Similar to standard PG, the details of these 11 practices were limited to definitions and lacked thorough step-by-step instructions. Providers in the PG condition attended a 1-day training event, with half a day covering study procedures. During the PG training, practices were described to the MHPs, with an opportunity for questions and answers. MHPs in this condition received no training on the multidimensional engagement framework nor the optimal pairing of practices with specific issues (i.e., which practices to use for which purpose). With these two enhancements to routine care, MHPs in the PG group knew that there were engagement concerns with the youth and family and also knew of potentially relevant procedures. As with the CKS group, no follow-up clinical consultation was provided throughout the study. In the study conclusion, providers in the PG condition were offered training in the RFES.

Measures

MTT: Youth (Y) and Caregiver (CG) Versions

We used two instruments to identify eligible youth cases with reported engagement concerns who would thus be the focus of clinical supervision and treatment for each MHP. The MTT-Y and MTT-CG are self-report instruments for youth and caregivers, consisting of 35 items, with seven items for each of the five dimensions of engagement (see Becker et al., 2018). Scales range from 0 to 21, with higher scores indicating higher engagement. Represented by the acronym REACH, scales include relationship (i.e., feeling understood and respected), expectancy (i.e., feeling that positive change is possible), attendance (i.e., perceptions of barriers to being present, on time, and prepared), clarity (i.e., appreciating how and why treatment should work), and homework (i.e., participating in exercises to be rehearsed within or outside of therapy sessions). The MTT-Y and MTT-CG are available in Spanish and English (see https://www.childfirst.ucla.edu/resources/), and research has shown that the REACH structure is invariant across ethnicity, language, and respondents (Chorpita & Becker, 2022). Cronbach's α values in the present study ranged from acceptable to good: relationship $\alpha_Y = .80$, $\alpha_{CG} = .93$; expectancy $\alpha_Y = .77$, $\alpha_{CG} = .91$; attendance $\alpha_Y = .60$, $\alpha_{CG} = .82$; clarity $\alpha_Y = .86$, $\alpha_{CG} = .95$; and homework $\alpha_Y = .79$, $\alpha_{CG} = .93$.

Action Cycle and Use of Evidence Behavioral Observation Coding System

We used the Action Cycle and Use of Evidence Behavioral Observation Coding System (ACEBOCS) as our measure of clinical reasoning and action that we examined in our primary aims. We coded transcribed recordings using the ACEBOCS to measure the activities, practices, and targets represented in 430 supervision and 208 treatment events. The ACEBOCS represents activities separate from the content of those activities (e.g., if a supervisory team considers using psychoeducation, the activity code = considers, and the content code = psychoeducation). For this study, we examined five activity codes (considers, selects, prepares, applies, and measures) and 21 content codes (i.e., seven problem codes, which included the five REACH domains plus nonengagement problem and other engagement problem, along with 14 practice codes: accessibility promotion; addressing barriers to treatment; appointment reminders; assessment; facilitating skill mastery; goal setting; motivational enhancement; positive expectation setting; psychoeducation about services; psychoeducation about the problem; relationship/rapport building; support networking; understanding identities, beliefs, and values; and other practice).

Coder training first involved a didactic review of coding procedures and codebook definitions, along with activities designed to help coders recognize codes within supervision transcripts. Subsequently, coders independently coded a transcribed supervision session that had been coded by the training team. Coders were trained to a criterion of reaching 80% or higher agreement on excerpt-level codes and event-level extensiveness ratings for two consecutive supervision events from a series of events that had been coded by a "gold standard" coding team. A similar training process was followed for treatment events, which were coded separately from supervision events and randomly assigned to coders. Following certification, coders attended weekly meetings to address questions about codes and reduce coder drift. Interrater reliability analyses conducted on a subsample of 84 supervision events and 78 treatment events from this trial indicated that coders achieved moderate to good reliability for ACEBOCS codes (e.g., median Fleiss' $\kappa = .68$ for activity codes; .59 for content codes; Becker et al., 2024). Descriptions of the full code system are available elsewhere (e.g., see Becker et al., 2024; Park et al., 2020).

Unified Theory of Acceptance and Use of Technology–2

We administered a 10-item adaptation of the Unified Theory of Acceptance and Use of Technology–2 (UTAUT-2) questionnaire (see Ahuna et al., 2023), based on the expanded unified theory of acceptance and use of technology model (Venkatesh et al., 2012). This measure was administered to MHPs at the end of the study to assess secondary outcomes related to perceived effort and utility in each study condition, given their relevance to the sustainability of the intervention approaches. The effort expectancy scale and performance

expectancy scale each contain three items, for which higher scores represent more positive perceptions. Items use a 7-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), and these scales have demonstrated good internal consistency (effort $\alpha = .93$; performance $\alpha = .94$; Ahuna et al., 2023).

Workload

Finally, as part of our secondary aim, we measured effort across conditions using both the duration of recorded supervision events and the sum of spoken words within each event. In our sample, word count and supervision duration were highly correlated ($r = .97$, $p < .001$). Word count was examined for the total supervision event and separately for the independent contributions of the therapist and supervisor.

Study Procedures

MTTs were administered to youth and caregivers from January 2018 to March 2020 (CA was completed in December 2019, prior to COVID-19 closures; participating SC schools remained open until March 2020). MTT administration for each youth case took place approximately 4–6 weeks after the start of their SMH services. MTT administration was fit to the existing procedures and preferences at each site, which typically involved administrative staff inviting youth or caregivers to complete the survey with other paperwork. All caregivers involved in treatment were eligible to be assessed. Our research team advised administrative staff that surveying children under the age of 7 might pose difficulties; thus, staff had the authority to use their judgment and provide reading assistance during the assessment. Due to local administrative procedures, a small number of youth cases (0.72%) were below age 7, which we chose to include in our study. CA surveys were mainly online, whereas most SC screening used paper-and-pencil format. At each site, administrators entered the therapist's name and youth's identification number before the youth and/or caregiver began the survey. Surveys completed online could not be seen by anyone other than the research team. After finishing paper surveys, participants placed their questionnaires in an envelope that they returned to the administrator, who then sent (e.g., by fax) questionnaires to the research team. Under no circumstances were completed surveys accessible to MHPs.

The research team scored all MTTs. Following the identification of an eligible case, the study team emailed the relevant MHPs. This email was delivered using an Health Insurance Portability and Accountability Act-secure email system using a standard message indicating that a particular youth was eligible for the study based on reported engagement concerns. For CKS participants, this email also included the engagement graphic with more detailed information about REACH subscale scores.

When notified, therapists obtained informed consent in the preferred language of each youth and caregiver at the next point of contact. Then, MHPs in both conditions digitally recorded three supervision events for consented youths and caregivers. Each supervision event took place in person in a dyadic format (i.e., one supervisor and one therapist). Days between supervision meetings had a median value of 14, with the distribution having a positive skew (e.g., 32% occurred within 1 week, 54% within 2 weeks, 71% within 3 weeks), and this distribution appeared to be driven by the frequency

of treatment sessions for this sample of students with low engagement (days between treatment sessions also had a median value of 14 days and were skewed similarly to supervision events; 32% occurred within 1 week, 53% within 2 weeks, 66% within 3 weeks). In addition, therapists recorded two treatment events following youth and caregiver study enrollment. Thus, the flow of digitally recorded supervision and therapy events within this study could be depicted as S_1-T_1-S_2-T_2-S_3. Recordings were collected, transcribed, independently checked for meeting transcription standards, and then independently double checked. Treatment conducted in Spanish was transcribed in Spanish, translated into English, and then independently double checked. Once suitable transcriptions were in place, they were coded using the ACEBOCS as described above. Participant study conditions and sites were masked. For the purposes of the present study, we selected records for only those events in the S_1-T_1-S_2 part of the sequence, to examine our primary hypotheses about the impact of the condition on supervision and the next immediate treatment session and supervision meeting, for which we had a priori expectations for defined activities to occur consistent with theory use of evidence in reasoning and treatment (e.g., Graham et al., 2006; Sackett et al., 2000). These recorded events represented over 200 hr of material involving 1,839,152 spoken words.

Data Analysis

Our primary analyses involved inspection of expected, model-consistent activities across the S_1-T_1-S_2 sequence, such that we examined whether the following planning and preparation codes occurred in the first supervision meeting (S_1): considers problem, selects problem, considers practice, selects practice, and prepares practice. We then inspected treatment (T_1) for the presence of an applies practice code. Finally, we inspected the second supervision event (S_2) for the presence of two monitoring codes: measures practice (i.e., reviewing how a practice was delivered) and measures problem (i.e., reviewing the impact on a youth or caregiver concern). Thus, all events were scored with binary outcomes for the presence or absence of the code. Activities in supervision events were indicated by at least one excerpt having that code assigned given the superior reliability of coding at the excerpt level (Becker et al., 2024). Treatment sessions were coded for practices using an extensiveness rating assigned by coders, and we assigned a positive score for treatment sessions when extensiveness ratings indicated a practice was clearly recognizable in the event (i.e., a 2 or above on a 0–5 scale, with 2 reflecting "more than a small part of the practice delivered" or "the practice was delivered with more than minimal depth/detail").

Because evidence-based reasoning and action were operationalized as a complex set of dependent activities, we deconstructed activities relevant to evidence-based practice according to three levels of focus. At the lowest level (any activity), we simply determined whether an activity occurred related to any problem or practice, a precondition for those activities to occur at any increased level of focus. For example, if a case had low engagement involving the relationship dimension, providers were credited with an activity being performed regardless of its relevance to the therapy relationship or even to engagement at all (e.g., considering the youth's trauma history in supervision would still yield a positive score for considers problem in this context; likewise, choosing to teach relaxation would still yield a positive

score for *selects practice*). At the next level (domain-focused), a positive score required that the problem code contained any relevance to engagement (whether or not it fit into the REACH framework) or that the practice code involved a procedure with known research support for any engagement problem. For example, for the youth screening positive for relationship concerns, a supervisor who discusses sporadic attendance (not related to the relationship) would nevertheless receive credit as having considers problem being domain-focused. At the third level, a subset of domain-focused activities were considered to be precision-focused if specific conditions were also met, consistent with a conceptualization of EBPP that involves joint consideration of the youth and family case-based evidence and concerns as well as the relevant research evidence. For example, for a youth who screened positive for relationship concerns, a supervisor who discusses behavioral indicators that the youth is uncomfortable or closed off during treatment (relevant to the relationship) would obtain credit for that event as having the considers problem activity being precision-focused. The specific criteria for performing activities at each of the three levels of focus appear in Table 5.

Given the nested nature of most of the data, we estimated mixed-effects regression models whenever appropriate. We considered up to four levels of nesting: events within cases within supervisees within supervisors. We first used null models to calculate the proportion of variance accounted for by each level of nesting and to explore model fit. When a level accounted for little variance in outcomes and did not significantly improve fit, we used a lower level model for parsimony. We generally estimated models with the following predictors: intercept, condition, site, and Condition × Site interaction. If the interaction term was not significant, we reran the model without the interaction term to estimate the main effects

of condition and state. Given the potential for large effects on some binary outcome variables (i.e., probabilities at or near 100%), mixed-effects regression models for the primary analyses were estimated using logistic regression models that applied a Firth-type penalty (Puhr et al., 2017). We calculated odds ratios (ORs) as the likelihood (expressed as odds) of an activity occurring in the CKS condition divided by the likelihood of that activity occurring in the PG condition. Larger ORs indicate a stronger association between the independent variable (condition) and the dependent variable (odds of an activity occurring) and thus serve as effect size indicators (Sánchez-Meca et al., 2003). We also calculated confidence intervals for all ORs, and statistically significant results indicate that the observed OR is different from 1.0, which represents no effect or equal odds (Szumilas, 2010). We chose an α level of .0001 for all logistic regression models given that we anticipated large effect sizes for most tests. All analyses were conducted in SAS 9.4, using PROC GLIMMIX for mixed-effects regression models with binary outcomes, PROC MIXED for mixed-effects regression models with continuous outcomes, and PROC LOGISTIC with Firth option for fixed-effects models. To aid in the pragmatic interpretation of these effects, we also report the observed probabilities of all outcomes across conditions in Figure 3.

Transparency and Openness

This study's design and analysis were not preregistered. The data sets generated and/or analyzed and the syntax needed to reproduce these analyses are not publicly available but are available from the corresponding author. These data have not been previously published.

Table 5
Definitions of Precision-Focused Practice by Activity and Content Type

Code	Requirement	Rationale
Considers problem, selects problem	Must be a dimension on which the case screened positive	Reflects start to planning that is relevant to the youth and family and reflects their preferences and considers case-based evidence
Considers practice, selects practice	Must be a procedure with supporting research evidence targeting (a) a dimension on which the case screened positive or (b) a dimension that the supervisor team explicitly chose to target	Reflects incorporating the best available research evidence pertaining to (a) case-based evidence or (b) an explicitly chosen target based on additional case material (i.e., not completely algorithmic based on measures, but reflective of case needs)
Prepares practice	Must be a procedure (a) with supporting research evidence targeting a dimension on which the case screened positive, (b) with supporting research evidence targeting a dimension that the supervisor team explicitly chose to target, (c) that the supervisory team explicitly considered for use, or (d) that the supervisory team explicitly chose to use in the next treatment session	Reflects (a) incorporating case-based and research evidence, (b) using research evidence relevant to an explicitly chosen target, (c) possible "trying on" of one or more practices for consideration by reviewing or rehearsing them first, or (d) preparing to deliver a practice explicitly chosen to fit the case and context
Applies practice	Must use a procedure with sufficient extensiveness to be considered complete (2 or higher on a 0–5 scale) that (a) was the practice selected in supervision or (b) had supporting research evidence targeting the engagement dimension which the supervisor team explicitly chose to target	Reflects (a) following the plan explicitly made in supervision or (b) pivoting to use a different practice supported by evidence given any circumstances in the immediate context (e.g., moving to a "plan B" consistent with the evidence base)
Measures practice, measures problem	Must be the procedure chosen or problem targeted explicitly by the supervisory team in the first supervision event (regardless of what happened in treatment)	Reflects integrity of plan and accountable practice (can allow for measurement of other problems and practices that emerged, but only credits what was chosen a priori)

Figure 3

Observed Rates of Occurrence of Selected Activities Across First Supervision, Treatment, and Second Supervision at Three Levels of Focus
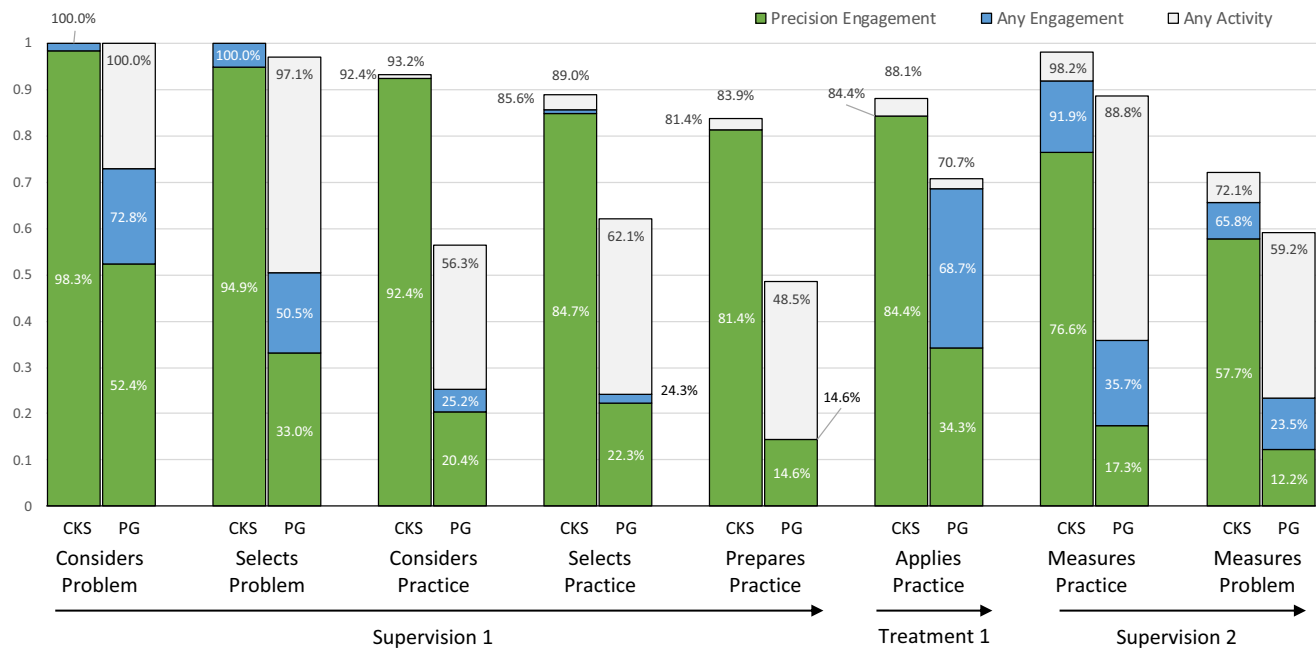
Note. CKS = coordinated knowledge system; PG = practice guidelines. See the online article for the color version of this figure.

Results

Primary Outcomes

Figure 3 shows the observed proportions of all primary implementation outcomes, as represented by eight activity-content combinations (e.g., considers problem, selects problem) measured across three observed events (S_1, T_1, or S_2) and presented by condition. For each outcome, observed proportions are presented for three levels of focus: any activity (i.e., activity occurred related to any problem or practice), domain-focused (i.e., activity occurred related to engagement), and precision-focused (i.e., activity occurred related to relevant case-based and research evidence).

The corresponding significance tests of differences in likelihood are based on ORs that are presented in Table 6. The eight activity-content combinations (e.g., considers problem, selects problem) are each reflected in one of the eight banded sections of the table. Within each activity-content combination, we present ORs for the three levels of focus (i.e., any activity, domain-focused, precision-focused) and compare the likelihood of each activity occurring across study conditions (CKS vs. PG) and across sites (CA vs. SC). No Site × Condition interaction effects emerged across any of the dependent variables for any level of focus. In terms of our hypotheses, these findings were consistent with the expectation that the effects of study condition were comparable across sites.

Any Activity

We first examined whether hypothesized differences would emerge even partially by examining whether relevant activities occurred at all, regardless of their relation to the research evidence or to any

engagement concerns. As shown in Figure 3, the considers problem code had a 100% rate of occurrence across both conditions (and sites), and thus, no analyses could be performed (i.e., for all cases everyone considered some problem of the youth and family). As presented in the any activity row for each of the eight outcomes in Table 6, significant main effects for study condition occurred only for considers practice, selects practice, and prepares practice, with ORs ranging from 4.72 to 10.39. These findings showed that the CKS group was significantly more likely in S_1 to engage in planning activities focused on any therapy procedures, with such activities occurring in more than 80% of the events in the CKS group versus just roughly over half of events in the PG group (see Figure 3, any activity).

Domain-Focused Activity

We then examined whether hypothesized differences would emerge in ways related to engagement, by examining whether activities occurred for which the problem codes were engagement related or the practice codes represented procedures with research supporting their effects for engagement problems. As shown in the domain-focused row for each of the eight outcomes in Table 6, significant main effects for study condition occurred for selects problem, considers practice, selects practice, prepares practice, measures practice, and measures problem, with ORs ranging from 6.10 to 297.28. These findings showed that the CKS group was significant more likely than the PG group in S_1 to select an engagement problem as a focus and to engage in planning activities focused on using procedures with research supporting their use for engagement, with such engagement practice-planning activities occurring in over 80% of supervision events in the CKS group versus roughly 25% or less in the PG group

Table 6
Supervision and Therapy Outcomes by Condition and Site

Activity type	Condition					Site				
	Penalized estimate	Penalized OR	95% CI	F	p	Penalized estimate	Penalized OR	95% CI	F	p
Considers problem										
Precision-focused	3.94	51.19	[15.93, 261.19]	32.84	<.0001*	1.63	5.11	[2.33, 11.78]	15.60	<.0001*
Domain-focused	4.74	114.00	[14.99, >999]	10.70	.0011	3.04	20.97	[6.19, 109.38]	18.55	<.0001*
Any activity[a]										
Selects problem										
Precision-focused	3.54	34.33	[15.04, 91.46]	60.29	<.0001*	0.44	1.55	[0.75, 3.27]	1.39	.2382
Domain-focused	5.69	297.28	[39.75, >999]	15.61	<.0001*	1.74	5.69	[2.51, 13.53]	16.43	<.0001*
Any activity	2.07	7.92	[0.76, >999]	2.08	.1494	0.56	1.74	[0.23, 19.47]	0.31	.5770
Considers practice										
Precision-focused	4.26	71.03	[28.22, 217.11]	68.70	<.0001*	1.86	6.43	[2.59, 18.86]	14.03	.0002
Domain-focused	3.99	53.91	[22.06, 155.80]	65.06	<.0001*	1.91	6.76	[2.86, 18.35]	16.67	<.0001*
Any activity	2.34	10.39	[4.82, 24.89]	31.74	<.0001*	1.00	2.73	[1.36, 5.60]	7.75	.0054
Selects practice										
Precision-focused	3.05	21.19	[10.59, 45.42]	67.52	<.0001*	1.17	3.22	[1.59, 6.85]	9.92	.0016
Domain-focused	3.02	20.54	[10.24, 44.20]	65.66	<.0001*	1.21	3.36	[1.66, 7.16]	10.69	.0011
Any activity	1.55	4.72	[2.41, 9.75]	18.97	<.0001*	0.66	1.94	[1.01, 3.78]	3.83	.0503
Prepares practice										
Precision-focused	3.62	37.27	[16.65, 95.20]	67.18	<.0001*	1.83	6.22	[2.77, 15.68]	17.50	<.0001*
Domain-focused	3.62	37.27	[16.65, 95.20]	67.18	<.0001*	1.83	6.22	[2.77, 15.68]	17.50	<.0001*
Any activity	1.74	5.71	[3.04, 11.12]	27.65	<.0001*	1.21	3.35	[1.79, 6.41]	13.76	.0002
Applies practice										
Precision-focused	2.48	11.96	[5.98, 25.57]	44.79	<.0001*	1.56	4.74	[2.38, 9.95]	18.23	<.0001*
Domain-focused	0.86	2.35	[1.19, 4.77]	5.86	.0155	1.45	4.26	[2.14, 8.94]	15.82	<.0001*
Any activity	1.06	2.89	[1.42, 6.16]	8.09	.0045	1.16	3.20	[1.57, 6.80]	9.75	.0018
Measures practice										
Precision-focused	3.25	25.66	[11.42, 65.71]	53.68	<.0001*	2.22	9.17	[4.08, 23.32]	25.29	<.0001*
Domain-focused	3.16	23.59	[10.60, 58.70]	52.81	<.0001*	1.50	4.49	[2.12, 10.13]	14.31	.0002
Any activity	1.70	5.46	[1.54, 28.86]	5.71	.0168	1.35	3.84	[1.18, 15.81]	4.48	.0344
Measures problem										
Precision-focused	2.23	9.28	[4.72, 19.49]	38.01	<.0001*	0.57	1.76	[0.93, 3.40]	2.91	.0879
Domain-focused	1.81	6.10	[3.35, 11.45]	33.08	<.0001*	0.69	1.99	[1.09, 3.70]	4.86	.0275
Any activity	0.55	1.73	[0.98, 3.10]	3.45	.0633	0.34	1.41	[0.79, 2.52]	1.36	.2439

Note. α level used for all tests was .0001. The first five activities were coded from S_1, and the final two were coded from S_2; applies practice results are shaded to indicate that they were obtained from T_1. OR = odds ratio; CI = confidence interval.
[a] Values could not be calculated for considers problem–any activity due to the absence of variance on this metric (all cells = 100%).
* $p < .0001$.

(see Figure 3, domain-focused). No significant differences were observed in T_1, with a moderate to high rate of the use of engagement practices in both groups, despite relatively minimal discussion and preparation of such practices in supervision within the PG group. In S_2, MHPs in the CKS condition were significantly more likely to discuss how an engagement practice was delivered (measures practice) and to review its impact on engagement (measures problem). Significant main effects for site emerged for considers problem, selects problem, considers practice, prepares practice, and applies practice, with ORs ranging from 4.26 to 20.97. These results suggested that in both study conditions, many of the activities in the first supervision and treatment events were more likely to focus on engagement among providers whose site was CA as compared with SC.

Precision-Focused Activity

Finally, we tested for differences that were consistent with our primary study aims of producing supervisory and treatment activity that incorporated both case-based and research evidence and was characterized by focused continuity across events. As shown in the precision-focused row for each of the eight outcomes in Table 6, significant main effects for study condition occurred for all activities measured in S_1, T_1, and S_2, with ORs ranging from 9.28 to 71.03. These findings showed that the CKS group was significantly more likely than the PG group to consider and select an engagement problem relevant to the youth and then to consider, select, and prepare to deliver procedures whose research evidence fit the problem on which the MHPs chose to focus. The OR was 11.96 for applies practice (i.e., providing the practice selected in supervision or using another engagement practice with supporting research evidence), which is a large effect on precision-focused use of case-relevant research evidence in treatment. Significant main effects for site emerged for considers problem, prepares practice, applies practice, and measures practice, with ORs ranging from 4.74 to 9.17. These results suggested that across both study conditions, various activities in both supervision and treatment were more likely to focus with greater precision on youth and family-relevant engagement among providers whose site was CA as compared with SC—most notably with respect to practice preparation (e.g., modeling, role play) prior to treatment and the review of those practices after treatment.

Secondary Outcomes

Performance Expectancy

UTAUT performance expectancy scores were first examined for supervisors, and a single-level analysis of variance model examining the main effects of study condition, site, and their interaction produced a significant main effect of study condition ($F = 5.93$; $p < .05$; $\eta^2 = .20$), with no significant site effect or Site $\times$ Condition interaction. Consistent with expectations, supervisors in the CKS condition rated their set of resources significantly more positively (i.e., more effective) than did supervisors in the PG condition (CKS $M = 18.54$, $SD = 2.37$; PG $M = 14.92$, $SD = 4.44$).

We then tested therapist UTAUT performance expectancy scores using a two-level mixed-effects model, with supervisees nested in supervisors (intraclass correlation coefficient [ICC] for the supervisor variance component was .23). This model showed a significant main effect of study condition ($F = 11.01$; $p < .05$), but

no significant site ($F = 3.93$; $p = .06$) or Site $\times$ Condition interaction ($F = 0.28$; $p = .60$). The pattern also supported our hypothesis, with therapists in the CKS condition rating their materials as significantly more effective than therapists in the PG condition (CKS $M = 16.86$, $SD = 2.41$; PG $M = 14.64$, $SD = 3.47$). Taken together, these results suggest that MHPs in the CKS condition perceived their resources as more effective in their work than did MHPs in the PG condition.

Effort Expectancy

UTAUT effort expectancy scores were then examined for supervisors, and a single-level analysis of variance model produced a significant main effect of study condition ($F = 8.21$; $p < .05$; $\eta^2 = .24$) with no significant site effect or Site $\times$ Condition interaction. Consistent with expectations, this effect demonstrated that supervisors in the CKS condition rated their set of resources significantly more positively (i.e., requiring less effort) than did supervisors in the PG condition (CKS $M = 18.31$, $SD = 2.02$; PG $M = 14.92$, $SD = 3.71$).

Therapist UTAUT responses for effort expectancy were then tested using the same mixed-effects model as for performance expectancy (ICC for the supervisor variance component was .29). This model showed significant main effects of study condition ($F = 8.26$; $p < .05$) and site ($F = 4.84$; $p < .05$), but no significant Site $\times$ Condition interaction ($F = 2.42$; $p = .14$). The pattern of results showed that therapists in the CKS condition rated their set of resources as requiring less effort than did therapists in the PG condition (CKS $M = 17.30$, $SD = 2.73$; PG $M = 15.13$, $SD = 3.79$), which supported our hypotheses. Regardless of condition, therapists in SC rated their resources as requiring less effort than did therapists in (CA $M = 15.44$, $SD = 3.86$; SC $M = 17.18$, $SD = 2.72$), which had not been predicted. Taken together, these results suggest that MHPs in the CKS condition perceived their resources as easier to use than did MHPs in the PG condition and that there were site differences, such that therapists in SC perceived all resources as easier to use than did therapists in CA.

Word Count

For total words spoken in supervision, we tested three-level mixed-effect models, with 430 supervision events nested within 92 therapists (ICC = .12) nested within 29 supervisors (ICC = .62). Tests for condition, site, and their interaction showed no effect of condition ($F = 0.40$, $p = .53$) and no interaction ($F = 0.68$, $p = .42$). There was, however, a significant effect of site ($F = 13.31$, $p < .05$), with SC teams having significantly lower word counts per supervision than CA teams, regardless of condition. Because event duration correlated almost perfectly with word count ($r = .97$), we did not perform these same tests for differences in the length of supervision. However, for context, the mean duration in minutes (SD) of supervision meetings by site and condition was as follows: CA|CKS = 20.39 (10.46), CA|PG = 20.00 (12.45), SC|CKS = 14.04 (6.85), and SC|PG = 6.23 (5.30).

We performed the same analysis for supervisor-spoken words, controlling for therapist-spoken words, to test whether the relative contribution of supervisors varied across conditions. There was no interaction ($F = 0.79$, $p = .38$) or site effect ($F = 0.84$, $p = .37$), but there was a main effect of condition ($F = 8.77$, $p < .05$). To put these findings in context, 37.0% ($SD = 13.6\%$) of spoken words were

attributable to supervisors in the PG condition, compared with 50.7% ($SD = 15.0\%$) in the CKS condition. Taken together, these results suggest that using the resources of the CKS did not significantly increase the observed effort of supervision above that put forth using the PG resource, as indicated by the total number of words spoken by the supervisor–therapist dyad. However, the CKS resources were associated with a significant shift in dyad dynamics, such that CKS supervisors spoke more than did PG supervisors.

Discussion

The use of a CKS considerably altered supervision and treatment in terms of its alignment with EBPP, most notably in terms of MHPs' tendency to incorporate case-specific and research evidence relevant to the youth or caregiver. The lack of Site × Condition interaction suggested further that the RFES was robust across different contexts and speaks to its generalizability. Importantly, these sites represented diverse workforces, who were mostly master's level professionals, the majority of whom were unlicensed, working in challenging contexts with high caseloads and demands common to community mental health. We consider these findings supportive of the notion that it is possible to improve clinical reasoning and URE in a highly representative workforce without the use of a manualized treatment (cf. Kazdin, 2019).

Operationalizing the process of activities and decisions as well as examining three levels of focus is a methodology that appears to hold promise for further specifying and experimentally testing evidence-based reasoning and action in clinical service. In particular, examining multiple levels of focus provides an ordinal view of high-quality evidence-based reasoning and URE (e.g., indicating when MHPs are "in the ballpark" but might lack a key resource in situ to act with greater precision), which lends itself logically to focal quality improvement of specific professional habits (e.g., improving the precise focus of consider practice activities). One case in point stems from the observed ubiquity of MHPs discussing problems in supervision. Regarding this activity, we infer that if a downstream aim is to apply a practice with some promise of addressing a chosen concern, it is not necessary to produce a "new" behavior in supervision but rather merely to improve its focus through a collection of supporting guides and materials. For preparation-related activities, on the other hand, which happened in less than half of supervision events in the PG condition, more support might be needed to produce robust habits of preparation.

The design and study outcomes suggest that changing the degree to which MHPs use evidence to guide their reasoning, planning, service delivery, and review need not require an intensive implementation effort, at least initially. The limited workshop training and lack of follow-up consultation run counter to what some research has shown to be most effective at changing practice (Beidas et al., 2012). That said, given that our measurement approach was to look for at least one instance of each indicated activity in supervision and in treatment, our results could reflect having merely initiated a process whereby MHPs were more informed. At present, we know less about the extensiveness or depth of these activities, which are the planned focus of additional research. Follow-up analyses can inspect the effects of condition on the depth and breadth of reasoning among MHPs, as well as how extensively any of the relevant practices were applied in treatment. Future trials could also test whether the activity

extensiveness benefits from ongoing consultation and feedback, which would be expected. At present, these results speak to how to reorganize the process of reasoning and action to incorporate case-based and research-based evidence in routine care.

These findings have similar implications for resource design, suggesting that there is value to building guidance systems that may not be full "treatments" or manuals but instead serve as prompts that structure collaborative interaction enriched with relevant research and case-based evidence (e.g., Chorpita & Daleiden, 2014). These CKSs should not be reified as manuals whose implementation requires "fidelity"; rather, we can imagine many ways to use the systems such as the RFES to improve various aspects of supervision and treatment, other than the way we approached it for the purposes of this study. Similarly, we believe that many such systems could be designed for various contexts, based on similar principles of informing and enriching decisions and actions in a dynamic, well-specified, and measurable workflow.

Site Differences

We intentionally chose two sites that differed significantly in terms of geographic region, resources, and organizational structure so that we could expose the CKS to a rigorous test in contexts that are representative of public mental health services. Differences were not observed at the "any" level of focus, suggesting that the context was not associated with these decisions or actions occurring in general. However, some domain-focused and precision-focused activities were more likely to occur in CA than in SC. There are a few possible explanations for this finding. MHPs in CA had a greater exposure to EBT trainings in general than did MHPs in SC. Thus, it is possible that MHPs in CA were more apt to apply something they had learned, perhaps because they had the benefit of earlier trainings they had attended or perhaps because the Los Angeles Unified School District organizational culture encouraged the supported application of what they had learned. Findings by Reeder et al. (2024), however, would suggest that were there an effect of training history, it would not be likely to improve MHP precision, which makes this explanation less plausible. On the other hand, it is possible that greater exposure to EBTs simply primes MHPs to be aware of some of the decisions and actions that are important when some type of interference (e.g., engagement concerns) arises during treatment. Another explanation is that it is possible that treatment engagement was more of an explicit priority within CA than within SC, perhaps due to accountability standards, such that MHPs attend more fully to it when concerns arise. Still, another possibility is that the significantly higher caseload demand in SC simply did not allow supervision to be as in depth (e.g., incorporating case material) as it was in CA.

Limitations and Constraints on Generalizability

There are numerous follow-up questions and hypotheses that we did not pursue in this first analysis of primary implementation outcomes for the Reaching Families trial. As noted above, future research should investigate the breadth of depth of various activities (e.g., how many practices were considered; how fully was each considered), whether and to what extent the sequence (and not merely the presence) of activities is important, including whether there are dependencies among activities (e.g., mediation). It would be helpful to examine the relative time spent on various activities

(e.g., percent of supervision), to examine whether such meta-skills as "activity transitions" (i.e., not spending too much time on one topic) are one of the mechanisms by which the RFES demonstrated its effects.

Although study condition and site were masked, ACEBOCS coders had general knowledge about the study design (e.g., that there were two conditions and two sites). It is possible that some coders inferred the study condition of a supervisor–therapist pair based on dyadic references to condition-specific resources or terminology. Whether correct or not, such guesses could bias coders to interpret transcribed materials according to their evolving presumptions rather than taking each passage as a novel material. For continued research on evidence-based reasoning and action using coding instruments such as the ACEBOCS, it will be important to test for reliability and bias in coding when there are minimal context cues (e.g., outside of a randomized controlled trial). Regarding generalizability constraints, again, we interpret the lack of Site × Condition effects along with the highly different sites selected for the study as evidence that these findings are likely to generalize to other community-based mental health systems serving diverse youth and whose workforce is a mix of primarily unlicensed master's level professionals representing minoritized backgrounds. Although we do not know whether the results would generalize to professionals outside the fields of social work and psychology who provide specialty mental health services to youth and families, our study sample is representative of the majority of those professionals who provide services within publicly funded mental health systems in the United States (e.g., Torpey, 2023).

Conclusions

Overall, these findings provide some assurances that with well-designed supports, MHPs in routine service contexts are likely to use case- and research-based evidence with some degree of precision in their supervision and service delivery, even with limited training. More generally, we see a promise for further investigation of evidence-based reasoning and action, as part of the maturing paradigms of EBPP and URE, when operationalized in a way that provides multiple levels of analysis (e.g., extent of focus) and that does not limit inspection of the service activity merely to what happens in treatment sessions (as is the case for the majority of measurement of treatment fidelity; Schoenwald et al., 2011). We also echo the sentiment of others (e.g., Weisz et al., 2018) that implementation efforts focused on supervision may be an efficient and comparatively scalable strategy to improve practice across a large workforce. In light of the present findings, we anticipate continued innovation and evaluation of diverse knowledge delivery tools and resources to inform and enrich the workflow of MHPs working in challenging contexts to improve the lives of youth and families.

References

Ahuna, J. K., Becker, K. D., & Chorpita, B. F. (2023). Predicting therapists' intentions to use innovations: Comparing the role of individual, organizational, and innovation characteristics. *Administration and Policy in Mental Health*, *50*(6), 946–965. https://doi.org/10.1007/s10488-023-01295-6

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

Barends, E., & Briner, R. (2014). Teaching evidence-based practice: Lessons from the pioneers: An interview with Amanda Burls and Gordon Guyatt. *Academy of Management Learning & Education*, *13*(3), 476–483. https://doi.org/10.5465/amle.2014.0136

Becker, K. D., Boustani, M., Gellatly, R., & Chorpita, B. F. (2018). Forty years of engagement research in children's mental health services: Multidimensional measurement and practice elements. *Journal of Clinical Child and Adolescent Psychology*, *47*(1), 1–23. https://doi.org/10.1080/15374416.2017.1326121

Becker, K. D., & Chorpita, B. F. (2023). Future directions in youth and family treatment engagement: Finishing the bridge between science and practice. *Journal of Clinical Child and Adolescent Psychology*, *52*(2), 284–309. https://doi.org/10.1080/15374416.2023.2169926

Becker, K. D., Park, A. L., Boustani, M. M., & Chorpita, B. F. (2019). A pilot study to examine the feasibility and acceptability of a coordinated intervention design to address treatment engagement challenges in school mental health services. *Journal of School Psychology*, *76*, 78–88. https://doi.org/10.1016/j.jsp.2019.07.013

Becker, K. D., Wu, E. G., Hukill, A., Brandt, N., & Chorpita, B. F. (2021). How do mental health providers assess treatment engagement of youth and caregivers? *Journal of Child and Family Studies*, *30*(10), 2527–2538. https://doi.org/10.1007/s10826-021-02042-x

Becker, K. D., Wu, E. G., Westman, J. G., Boyd, M. R., Guan, K., Lakind, D., Chu, W., Knudsen, K. S., Bradley, W. J., Park, A. L., Kenworthy LaMarca, T., Lang, E., & Chorpita, B. F. (2024). The interrater reliability of a coding system for measuring mental health professionals' decisions and actions. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication. https://doi.org/10.1080/15374416.2024.2384027

Beidas, R. S., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, *63*(7), 660–665. https://doi.org/10.1176/appi.ps.201100401

Campbell, M. K., Elbourne, D. R., Altman, D. G., & the CONSORT Group. (2004). CONSORT statement: Extension to cluster randomised trials. *The BMJ*, *328*(7441), Article 702. https://doi.org/10.1136/bmj.328.7441.702

Chambers, D. A., & Norton, W. E. (2016). The adaptome: Advancing the science of intervention adaptation. *American Journal of Preventive Medicine*, *51*(4), S124–S131. https://doi.org/10.1016/j.amepre.2016.05.011

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7–18. https://doi.org/10.1037/0022-006X.66.1.7

Chorpita, B. F., & Becker, K. D. (2022). Dimensions of treatment engagement among youth and caregivers: Structural validity of the REACH framework. *Journal of Consulting and Clinical Psychology*, *90*(3), 258–271. https://doi.org/10.1037/ccp0000711

Chorpita, B. F., & Daleiden, E. L. (2014). Structuring the collaboration of science and service in pursuit of a shared vision. *Journal of Clinical Child and Adolescent Psychology*, *43*(2), 323–338. https://doi.org/10.1080/15374416.2013.828297

Chorpita, B. F., & Daleiden, E. L. (2018). Coordinated strategic action: Aspiring to wisdom in mental health service systems. *Clinical Psychology: Science and Practice*, *25*(4), Article e12264. https://doi.org/10.1111/cpsp.12264

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., Phillips, L., Ward, A., Lynch, R., Trent, L., Smith, R. L., Okamura, K., & Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, *18*(2), 154–172. https://doi.org/10.1111/j.1468-2850.2011.01247.x

Crowley, D. M., Scott, J. T., Long, E. C., Green, L., Giray, C., Gay, B., Israel, A., Storace, R., McCauley, M., & Donovan, M. (2021). Cultivating researcher-policymaker partnerships: A randomized controlled trial of a model for training public psychologists. *American Psychologist*, *76*(8), 1307–1322. https://doi.org/10.1037/amp0000880

Deming, W. (1993). *The new economics*. MIT Press.

Djulbegovic, B., & Guyatt, G. H. (2017). Progress in evidence-based medicine: A quarter century on. *Lancet*, *390*(10092), 415–423. https://doi.org/10.1016/S0140-6736(16)31592-6

Duong, M. T., Bruns, E. J., Lee, K., Cox, S., Coifman, J., Mayworm, A., & Lyon, A. R. (2021). Rates of mental health service utilization by children and adolescents in schools and other common service settings: A systematic review and meta-analysis. *Administration and Policy in Mental Health*, *48*(3), 420–439. https://doi.org/10.1007/s10488-020-01080-9

Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvement targets by investigating usual mental health care. *Administration and Policy in Mental Health*, *37*(1–2), 15–26. https://doi.org/10.1007/s10488-010-0279-y

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: Time for a map? *The Journal of Continuing Education in the Health Professions*, *26*(1), 13–24. https://doi.org/10.1002/chp.47

Guan, K., Kim, R. E., Rodas, N. V., Brown, T. E., Gamarra, J. M., Krull, J. L., & Chorpita, B. F. (2019). Emergent life events: An in-depth investigation of characteristics and provider responses during youth evidence-based treatment. *Journal of Clinical Child and Adolescent Psychology*, *48*(6), 906–921. https://doi.org/10.1080/15374416.2018.1496441

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109–112. https://doi.org/10.4103/2229-3485.83221

Henggeler, S. W., & Schoenwald, S. K. (2002). Treatment manuals: Necessary, but far from sufficient. *Clinical Psychology: Science and Practice*, *9*(4), 419–420. https://doi.org/10.1093/clipsy.9.4.419

Kazdin, A. E. (2019). Annual Research Review: Expanding mental health services through novel models of intervention delivery. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *60*(4), 455–472. https://doi.org/10.1111/jcpp.12937

Kazdin, A. E. (2023). *Research design in clinical psychology* (6th ed.). Cambridge University Press. https://doi.org/10.1017/9781108976589

Kendall, P. C., & Chu, B. C. (2000). Retrospective self-reports of therapist flexibility in a manual-based treatment for youths with anxiety disorders. *Journal of Clinical Child Psychology*, *29*(2), 209–220. https://doi.org/10.1207/S15374424jccp2902_7

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, *26*(1), 87–119. https://doi.org/10.1145/174666.174668

Office of Human Rights and the Mayor's Office of Racial Equity. (2022, April). *Guide to inclusive language: Race and ethnicity*. https://ohr.dc.gov/sites/default/files/dc/sites/ohr/page_content/attachments/OHR_ILG_RaceEthnicity_FINAL%20%281%29.pdf

Park, A. L., Becker, K. D., Boustani, M. M., & Chorpita, B. F. (2020). Decision-making in youth mental health care: Measuring provider and supervisor use of evidence. *Administration and Policy in Mental Health*, *47*(3), 344–356. https://doi.org/10.1007/s10488-019-00989-0

Pellerin, K. A., Costa, N. M., Weems, C. F., & Dalton, R. F. (2010). An examination of treatment completers and non-completers at a child and adolescent community mental health clinic. *Community Mental Health Journal*, *46*(3), 273–281. https://doi.org/10.1007/s10597-009-9285-5

Puhr, R., Heinze, G., Nold, M., Lusa, L., & Geroldinger, A. (2017). Firth's logistic regression with rare events: Accurate effect estimates and predictions? *Statistics in Medicine*, *36*(14), 2302–2317. https://doi.org/10.1002/sim.7273

Real, K., & Poole, M. S. (2005). Innovation implementation: Conceptualization and measurement in organizational research. *Research in Organizational Change and Development*, *15*, 63–134. https://doi.org/10.1016/S0897-3016(04)15003-9

Reeder, K., Becker, K. D., Chu, W., & Chorpita, B. F. (2024). Characterizing provider readiness for meeting engagement challenges in youth psychotherapy. *Evidence-Based Practice in Child and Adolescent Mental Health*, *9*(3), 332–343. https://doi.org/10.1080/23794925.2024.2330398

Regan, J., Lau, A. S., Barnett, M., Stadnick, N., Hamilton, A., Pesanti, K., Bando, L., & Brookman-Frazee, L. (2017). Agency responses to a system-driven implementation of multiple evidence-based practices in children's mental health services. *BMC Health Services Research*, *17*(1), Article 671. https://doi.org/10.1186/s12913-017-2613-5

Rickinson, M., Cirkony, C., Walsh, L., Gleeson, J., Cutler, B., & Salisbury, M. (2022). A framework for understanding the quality of evidence use in education. *Educational Research*, *64*(2), 133–158. https://doi.org/10.1080/00131881.2022.2054452

Rotheram-Borus, M. J., Swendeman, D., & Chorpita, B. F. (2012). Disruptive innovations for designing and diffusing evidence-based interventions. *American Psychologist*, *67*(6), 463–476. https://doi.org/10.1037/a0028180

Rousseau, D. M., & Gunia, B. C. (2016). Evidence-based practice: The psychology of EBPP implementation. *Annual Review of Psychology*, *67*(1), 667–692. https://doi.org/10.1146/annurev-psych-122414-033336

Sackett, D. L. (1981). How to read clinical journals: I. why to read them and how to start reading them critically. *Canadian Medical Association Journal*, *124*(5), 555–558.

Sackett, D. L. (1989). Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*, *95*(2), 2S–4S. https://doi.org/10.1378/chest.95.2_Supplement.2S

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *The BMJ*, *312*(7023), Article 71. https://doi.org/10.1136/bmj.312.7023.71

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM*. Churchill Livingstone.

Saloner, B., Carson, N., & Cook, B. L. (2014). Episodes of mental health treatment among a nationally representative sample of children and adolescents. *Medical Care Research and Review: MCRR*, *71*(3), 261–279. https://doi.org/10.1177/1077558713518347

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448

Schoenwald, S. K., Garland, A. F., Southam-Gerow, M. A., Chorpita, B. F., & Chapman, J. E. (2011). Adherence measurement in treatments for disruptive behavior disorders: Pursuing clear vision through varied lenses. *Clinical Psychology: Science and Practice*, *18*(4), 331–341. https://doi.org/10.1111/j.1468-2850.2011.01264.x

Southam-Gerow, M. A., Chorpita, B. F., Miller, L. M., & Gleacher, A. A. (2008). Are children with anxiety disorders privately referred to a university clinic like those referred from the public mental health system? *Administration and Policy in Mental Health*, *35*(3), 168–180. https://doi.org/10.1007/s10488-007-0154-7

Supplee, L., Boaz, A., & Metz, A. (2023). *Learning across contexts: Bringing together research on research use and implementation science*. William T. Grant Foundation. https://wtgrantfoundation.org/wp-content/uploads/2023/02/Supplee-Boaz-Metz_WTG-Digest-8.pdf

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227–229.

Thorne, F. C. (1947). The clinical method in science. *American Psychologist*, *2*(5), 159–166. https://doi.org/10.1037/h0060157

Torpey, E. (2023, October). *Careers in mental health services. Career outlook*. U.S. Bureau of Labor Statistics. https://www.bls.gov/careeroutlook/2023/article/careers-in-mental-health-services.htm#:~:Text=The%20occupations%20in%20the%20table,for%20all%20occupations%20in%202022

Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *Management Information Systems Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Weisz, J. R., Ugueto, A. M., Herren, J., Marchette, L. K., Bearman, S. K., Lee, E. H., Thomassin, K., Alleyne, A., Cheron, D. M., Tweed, J. L., Hersh, J., Raftery-Helmer, J. N., Weissman, A. S., & Jensen-Doss, A. (2018). When the torch is passed, does the flame still burn? Testing a "train the supervisor" model for the Child STEPs treatment program. *Journal of Consulting and Clinical Psychology*, *86*(9), 726–737. https://doi.org/10.1037/ccp0000331

Weisz, J. R., Venturo-Conerly, K. E., Fitzpatrick, O. M., Frederick, J. A., & Ng, M. Y. (2023). What four decades of meta-analysis have taught us about youth psychotherapy and the science of research synthesis. *Annual Review of Clinical Psychology*, *19*(1), 79–105. https://doi.org/10.1146/annurev-clinpsy-080921-082920

Westmacott, R., & Hunsley, J. (2017). Psychologists' perspectives on therapy termination and the use of therapy engagement/retention strategies. *Clinical Psychology & Psychotherapy*, *24*(3), 687–696. https://doi.org/10.1002/cpp.2037

Wiltsey-Stirman, S., & Comer, J. S. (2018). What are we even trying to implement? Considering the relative merits of promoting evidence-based protocols, principles, practices, or policies. *Clinical Psychology: Science and Practice*, *25*(4), Article e12269. https://doi.org/10.1111/cpsp.12269

Wiltsey-Stirman, S., Gamarra, J., Bartlett, B., Calloway, A., & Gutner, C. (2017). Empirical examinations of modifications and adaptations to evidence-based psychotherapies: Methodologies, impact, and future directions. *Clinical Psychology: Science and Practice*, *24*(4), 396–420. https://doi.org/10.1037/h0101747

Witmer, L. (1996). Clinical psychology. *American Psychologist*, *51*(3), 248–251. https://doi.org/10.1037/0003-066X.51.3.248 (Original work published 1907)

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at https://my.apa.org/portal/alerts/ and you will be notified by e-mail when issues of interest to you become available!