Stakeholder intention to engage in fidelity measurement methods in community mental health settings: A mixed methods study

Carlin P. Hoffacker1,2, Melanie Klein2, Emily M. Becker-Haimes2,3, Jessica Fishman2,4,5, Sonja K. Schoenwald6, Perrin B. Fugo2, Bryce D. McLeod7, Shannon Dorsey4,8, Shannon Litke9, Lah’Nasia Shider2, Adina Lieberman2, David S. Mandell2, and Rinad S. Beidas2,4,10,11,12,13,14

Abstract

Background

The current gold standard for measuring fidelity (specifically, adherence) to cognitive behavioral therapy (CBT) is direct observation, a costly, resource-intensive practice that is not feasible for many community organizations to implement regularly. Recent research indicates that behavioral rehearsal (i.e., role-play between clinician and individual with regard to session delivery) and chart-stimulated recall (i.e., brief structured interview between clinician and individual about what they did in session; clinicians use the client chart to prompt memory) may provide accurate and affordable alternatives for measuring adherence to CBT in such settings, with behavioral rehearsal yielding greater correspondence with direct observation.

Methods

Drawing on established causal theories from social psychology and leading implementation science frameworks, this study evaluates stakeholders’ intention to use behavioral rehearsal and chart-stimulated recall. Specifically, we measured attitudes, self-efficacy, and subjective norms toward using each, and compared these factors across the two methods. We also examined the relationship between attitudes, self-efficacy, subjective norms, and intention to use each method. Finally, using an integrated approach we asked stakeholders to discuss their perception of contextual

1Department of Counseling and Educational Psychology, Indiana University, Bloomington, IN, USA
2Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA
3Hall Mercer Community Mental Health, University of Pennsylvania Health System, Philadelphia, PA, USA
4Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, USA
5Message Effects Lab, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA
6Oregon Social Learning Center, Eugene, OR, USA
7Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA
8Department of Psychology, University of Washington, Seattle, WA, USA
9Department of Psychology, Drexel University, Philadelphia, PA, USA
10Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA, USA
11Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA
12Penn Implementation Science Center at the Leonard Davis Institute (PISCE@LDI), University of Pennsylvania, Philadelphia, PA, USA
13Penn Medicine Nudge Unit, University of Pennsylvania Health System, Philadelphia, PA, USA
14Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Corresponding author:

Rinad S. Beidas, 325 N. Michigan Avenue, 21st Floor, Chicago, IL 60661, USA.
Email: rinad.beidas@northwestern.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
factors that may influence beliefs about using each method. These data were collected from community-based supervisors \((n = 17)\) and clinicians \((n = 66)\).

**Results**

Quantitative analyses suggest moderately strong intention to use both methods across stakeholders. There were no differences in supervisors’ or clinicians’ attitudes, self-efficacy, subjective norms, or intention across methods. More positive attitudes and greater reported subjective norms were associated with greater reported intention to use either measure. Qualitative analyses identified participants’ specific beliefs about using each fidelity measure in their organization, and results were organized using the Consolidated Framework for Implementation Research.

**Conclusions**

Strategies are warranted to overcome or minimize potential barriers to using fidelity measurement methods and to further increase the strength of intention to use them.

**Plain Language Summary:** The best way to measure fidelity, or how closely a clinician follows the protocol, to Cognitive Behavioral Therapy (CBT) is watching the session. This is an expensive practice that is not feasible for many community organizations to do regularly. Recent research indicates that behavioral rehearsal, or a role-play between the clinician and individual with regard to session delivery, and chart-stimulated recall, or a brief discussion between an individual and the clinician about what they did in session with the clinician having access to the chart to help them remember, may provide accurate and affordable alternatives for measuring fidelity to CBT. We just completed a study demonstrating that both methods are promising, with behavioral rehearsal offering scores that are the most similar to watching the session. Drawing on established theories from social psychology and leading implementation science frameworks, this study evaluates future supervisor and clinician motivation to use these fidelity measurement methods. Specifically, we measured supervisor \((n = 17)\) and clinician \((n = 66)\) attitudes, norms, self-efficacy, intentions, and anticipated barriers and facilitators to using each of these fidelity measurement tools. Quantitative and qualitative analyses suggest similar intention to use both methods, and concerns about barriers to using each method. Further research is warranted to minimize the burden associated with implementing fidelity measurement methods and deploying strategies to increase use.

**Keywords**

community mental health, fidelity measurement, behavioral intention, mixed methods

Implementation research and quality improvement efforts require accurate measurement of the fidelity of to psycho-social interventions delivered in community settings (England et al., 2015). Fidelity is (1) often hypothesized as the mechanism through which client outcomes improve; (2) the most proximal indicator of treatment quality (McLeod et al., 2013); and (3) the primary implementation outcome of interest (Proctor et al., 2011). Until recently, accurate and low-cost methods to measure fidelity, which include both adherence to the components of an intervention and the skillfulness with which the intervention is executed (i.e., competence), did not exist (Cox et al., 2019; Perepletchkova et al., 2007). Direct observation, the gold standard for assessing fidelity of psycho-social interventions, is resource-intensive and nearly impossible to implement in community settings.

A recently conducted randomized controlled trial (RCT) tested novel methods to assess fidelity (specifically, adherence, or how extensively a clinician follows a protocol as intended) to cognitive-behavioral therapy (CBT) delivered to pediatric populations compared with direct observation. Two promising methods emerged: behavioral rehearsal (i.e., role-play between clinician and individual) (Beidas et al., 2014) and chart-stimulated recall (CSR) (i.e., brief structured interview between an individual and the clinician about what they did in session; clinicians use the chart to prompt their memory) (Cunnington et al., 1997; Goulet et al., 2007; Jennett & Affleck, 1998; Miller et al., 2010; Salvatori et al., 2008). Behavioral rehearsal (BR) was most accurate, but CSR also performed well. These findings suggest the potential for these two methods to be used in lieu of direct observation in implementation research and practice, as they require fewer resources (Becker-Haimes et al., 2020; Becker-Haimes et al., 2022). Economic evaluation of the two methods is forthcoming. The current study was designed to identify whether one has a greater chance of being used in community mental health settings and why.

Causal models from social psychology, such as the Theory of Planned Behavior (TPB), can help predict how likely one is to use a method and explain why. In these models, behavioral intention is the most proximal determinant of behavior (Fishbein & Ajzen, 2010; Sheeran, 2002) and can therefore help explain and predict the implementation of evidence-based practices (Fishman et al., 2020). In the present study, behavioral intention refers to the amount of effort one is willing to expend to implement
BR or CSR. To use either method, supervisors and clinicians must have a relatively strong intention to do so. The strength of intention to use a fidelity measurement method may depend on the method, the people using it, and the context in which they work (Fishman et al., 2018; Wolk et al., 2019). Given that supervisors and clinicians typically are not required to use BR or CSR, relatively strong intention to use a method will increase the likelihood it is used (Eccles et al., 2007; Webb & Sheeran, 2006).

This study examines factors explicitly posited in the TPB. Behavioral intention is a function of one’s attitudes, perceived behavioral control, (also referred to as self-efficacy), and subjective norms (Fishbein & Ajzen, 2010; Sheeran, 2002; Webb & Sheeran, 2006). Attitudes are defined as the perceived advantages or disadvantages of using BR or CSR. Self-efficacy is a construct representing one’s perceived ability to perform a behavior, such as routinely using BR or CSR. Subjective norms include the perception of whether important others expect one to use BR or CSR and also the perception of whether other clinicians are using these methods. These three factors are malleable and can be targeted by strategies when intention to engage in a specific behavior (in this case, use of BR or CSR) is relatively weak (Fishbein & Ajzen, 2010).

In addition to examining psychological factors per the TPB, this study also examines contextual factors often studied and prioritized in implementation science frameworks (e.g., organizational culture and climate). Contextual factors, drawn from the Consolidated Framework for Implementation Research (CFIR) (Damschroder et al., 2009), are important to consider because stakeholders’ perceptions of contextual factors can influence their beliefs about using a method and also moderate the relationship between intentions and behavior (Fishman et al., 2018, 2020). Even for those who are motivated to use a fidelity measurement method, their context must make it possible (and ideally easy) to follow through. Policies, resources, and other organizational characteristics can influence the likelihood that one is able to use BR or CSR when motivated to do so (Fishman et al., 2018, 2020). For example, a clinician who is highly motivated to use a method regularly but lacks the needed materials is unlikely to use that method frequently.

In this mixed-methods study, we endeavored to address the following aims: (1) Estimate the strength of intention to use BR and CSR in a sample of community mental health clinicians experienced with the methods (2) Measure attitudes, self-efficacy and subjective norms, which explain the strength of intention; and (3) Identify contextual factors from the CFIR, such as organizational culture and climate, that may underly these attitudes, self-efficacy, and subjective norms. We hypothesized that supervisors and clinicians would have stronger intention and perceive fewer contextual factors that might serve to impede the use of CSR compared with BR because the former most closely resembles existing supervisory processes in community organizations (Dorsey et al., 2018) (Beidas et al., 2014) and may be perceived as less anxiety provoking (Edmunds et al., 2013).

Methods

The City of Philadelphia and the University of Pennsylvania Institutional Review Boards approved this study. Informed written or verbal consent or assent was obtained from all participants.

Study Setting and Agency Participation

Using a list of organizations accepting public mental health insurance within a 50-mile radius of Philadelphia County, we contacted agency leadership via phone and email to invite them to participate in the RCT (Becker-Haimes et al., 2022). Outpatient mental health agencies or residential treatment facilities were eligible if they provided care to youth clients (ages 7–24) and employed clinicians who used at least one of 12 CBT techniques measured for the study (e.g. operant strategies targeted at youth). Supervisor, clinician, and client participants (youth and, in the case of minors, their guardians) were recruited from 27 community-based agencies providing mental health services to youth in the Philadelphia tri-state area.

Leadership from each agency invited clinicians to attend an informational meeting conducted at each agency. Research staff introduced the study to clinicians at these meetings and enrolled interested parties. Clinicians were eligible if they endorsed having at least three youth (7–24) clients on their caseload with whom they planned to use at least one CBT technique in the next month. Participants completed quantitative and qualitative evaluations. No supplementary relationships were established or preexisted between interviewers and participants. No specific information about interviewers was formally disclosed aside from affiliation with the University of Pennsylvania and standard informed consent requirements (e.g., mandated reporting).

Procedures and Sampling

Qualitative Interview Procedures

Qualitative interviews were conducted by trained, bachelor- or master-level research coordinators (including authors Hoffacker, Fugo, and Lieberman). Pilot interviews were conducted by PhD-level research team members (Drs. Becker-Haimes and Beidas). All interviewers were women. Interviewers were trained by reviewing previous interviews, shadowing trained interviewers, conducting mock interviews, and conducting two supervised interviews. Participants completed interviews in a private space at their place of work. Qualitative interviews lasted roughly one hour. No repeat interviews were conducted, transcripts were not returned to participants for comment.
or correction, and participants did not provide feedback on findings. Interview themes converged and reached saturation, and additional interviews were not needed.

**Supervisors (n = 17)**

One supervisor per agency was invited to complete a structured questionnaire that measured strength of attitudes, self-efficacy, subjective norms, and intention to use the measurement methods, and to participate in a qualitative semi-structured interview. Ten supervisors declined to participate in either portion of the study due to schedules. Fourteen supervisors completed the questionnaire and 17 completed the interview. Due to time constraints, three supervisors who completed the interview did not complete the quantitative questionnaire. Supervisors were compensated $50 per hour.

**Clinicians (n = 66)**

Of 218 eligible clinicians invited to participate in the larger study, 84 declined to participate, eight listed reasons other than preference not to participate and ineligibility (e.g., leadership requested that interns not participate), yielding 126 clinicians randomly assigned to study conditions. Seven clinicians withdrew, five discontinued due to failure to enroll any clients, and eight were unable to complete study activities due to interruption caused by coronavirus disease 2019 (COVID-19). Two clinicians who completed all study activities virtually due to COVID-19 were excluded from analyses, as was one clinician who completed the BR condition during piloting (after which the model of the measure changed meaningfully), yielding 103 clinicians who were included in quantitative analysis. Importantly, only clinicians assigned to BR (n = 32) and CSR (n = 34) conditions are included in the present study, as these conditions of the RCT were most promising with regard to their accuracy (Becker-Haines et al., 2020; Becker-Haines et al., 2022).

Clinicians completed a structured questionnaire that measured attitudes, self-efficacy, subjective norms (referred to as norms going forward), and intention after completing the measurement method to which they were assigned in the RCT. Thirty-six randomly selected participating clinicians (roughly one from each condition per agency) were invited to participate in an additional qualitative semi-structured interview upon completion of their measurement condition to assess attitudes, self-efficacy, norms, and contextual factors. One clinician in the BR condition was excluded from qualitative analyses as all study activities were conducted virtually due to COVID-19. The dataset reported here consists of only clinicians randomized to the BR (n = 17) and CSR (n = 18) conditions of the RCT. All clinicians invited to complete the additional interview agreed. Clinicians were compensated $50 per hour.

Qualitative interviews were recorded using a digital audio recorder, inputted into REDCap (Harris et al., 2019), transcribed, and coded using NVivo12 software and Microsoft Excel.

**Measures**

**TPB Constructs**

Consistent with our first and second aim, we examined TPB constructs in two ways. First, among supervisors and clinicians, we used gold standard established quantitative methods to measure attitudes, norms, self-efficacy, and intention to use each of the two methods (Fishman et al., 2020). Scales were modeled from items developed by Fishbein & Azjen (2010), typically yielding alphas >0.80 and test-retest reliability of ≥0.70. Examples of items from the clinician version are below; parallel items were developed for supervisors. Supervisor questionnaires inquired about each method included in the study while clinician questionnaires inquired only about the condition to which they were randomized.

Second, we followed qualitative belief elicitation interview procedures associated with the TPB and similar causal models. Where quantitative measures offered a measure of the directionality of constructs underlying intention (e.g., how positive or negative attitudes are), qualitative measures assess the specific beliefs underlying positive or negative constructs (Erbe et al., 2020; Lederer & Middlestadt, 2014).

**TPB Construct Quantitative Questionnaires**

**Attitudes.** Attitudes towards each method were measured by six items asking respondents to use scales anchored by opposite adjective pairs to evaluate their use of this method as relatively “burdensome” versus “manageable,” “foolish” versus “wise,” “bad” versus “good,” “useless” versus “effective,” “confusing” versus “clear,” “unpleasant” versus “pleasant.” The response options, which ranged from 0 to 10, were averaged across these items.

**Self-efficacy.** Self-efficacy was measured with two items (response range 1 “strongly disagree” to 5 “strongly agree”): “It is totally up to me whether I participate in [BR/CSR],” (to measure perceived behavioral control) and “Participating in [BR/CSR] is something I could learn how to do.” The overall score represents an average of these two items.

**Subjective Norms.** Norms were measured with two items (response range 1 “strongly disagree” to 5 “strongly agree”): “Most clinicians would approve of me participating in [BR/CSR],” “Most clinicians with a job like mine would participate in [BR/CSR].” The overall score represents an average of these two items.

**Intention.** Intention was measured using one item (response range 1 “very unlikely” to 7 “very likely”): “Imagine that your supervisor uses behavioral rehearsal/chart-stimulated recall about once a month with you to
evaluate your fidelity to CBT. How likely are you to participate in this type of supervision?"

**Demographics**

Demographics included age, gender, race/ethnicity, licensure status, education, role within the agency, human services work history, weekly hours spent working at the present agency, and average time allotted for supervision (Table 1). Clinicians were also asked about the typical caseload, year they initiated CBT in their work, and participation in city-wide training initiatives in evidence-based practices.

**Qualitative Interviews to Elicit Beliefs Informing Attitudes, Self-Efficacy, and Norms and Contextual Factors**

The qualitative interviews had two goals: (1) to identify specific beliefs informing attitudes, self-efficacy, and subjective norms; and (2) to understand contextual factors that may underly these beliefs or moderate the relationship between intention and behavior. Figure 1 depicts the relationships between these constructs via three pathways: Pathway A, the influence of contextual factors on the TPB constructs of attitudes, self-efficacy, and subjective norms; Pathway B, TPB constructs informing behavioral intention; and Pathway C, moderating effect of contextual factors on the intention-behavior gap. In the interviews, we focused on Pathway A because these methods have not yet been adopted in routine care in community mental health. Semi-structured individual interviews with clinicians were conducted using 10 open-ended questions to identify salient “tip of the tongue” beliefs that spontaneously come to mind when thinking about the behavior of interest (Sutton et al., 2003). Rather than relying on preconceived lists of statements identified by researchers, this method identifies beliefs commonly reported by the population of interest.

To identify beliefs underlying attitudes, we asked respondents to share the perceived advantages and disadvantages of using BR and/or CSR. To identify beliefs underlying self-efficacy, we asked what would make it difficult or easier to routinely use BR and/or CSR. To identify beliefs underlying norms, we asked participants who would approve or disapprove of them routinely using BR and/or CSR. The interview guide was rehearsed with members of the research team, and three pilot interviews

---

**Table 1. Supervisor and clinician demographics.**

| Variable                  | Supervisor n = 17 | Clinician BR n = 32 | Clinician CSR n = 34 |
|---------------------------|-------------------|---------------------|----------------------|
| Female gender             | 10 (58.8)         | 28 (87.5)           | 27 (79.4)            |
| Hispanic/Latinx           | 3 (17.6)          | 4 (12.5)            | 2 (5.9)              |
| Racial identity           |                   |                     |                      |
| White/Caucasian           | 14 (82.4)         | 22 (68.8)           | 29 (85.3)            |
| Asian or Pacific Islander | 0 (0)             | 3 (9.4)             | 1 (2.9)              |
| Black or African American | 2 (11.8)          | 5 (15.6)            | 4 (11.8)             |
| Other                     | 1 (5.9)           | 1 (3.1)             | 0 (0)                |
| Age (M, SD)               | 40.18 (10.48)     | 36.35 (12.62)       | 37.71 (12.43)        |
| Education                 |                   |                     |                      |
| Bachelor’s degree         | –                 | 1 (3.1)             | 1 (2.9)              |
| Master’s degree           | 14 (82.4)         | 29 (90.6)           | 33 (97.1)            |
| Doctorate                 | 2 (11.8)          | 2 (6.3)             | 1 (2.9)              |
| Type of employee          |                   |                     |                      |
| Salaried                  | –                 | 17 (53.1)           | 17 (50)              |
| Fee-for-service           | –                 | 13 (40.6)           | 17 (50)              |
| M (SD)                    | –                 | –                   | –                    |
| Years of experience       |                   |                     |                      |
| In human services         | 15.73 (11.39)     | 9.28 (10.95)        | 10.24 (9.99)         |
| As a supervisor           | 6.22 (3.92)       | –                   | –                    |
| Hours of supervision per week |            |                     |                      |
| Individual supervision    | –                 | 0.83 (0.48)         | 0.88 (0.50)          |
| Group supervision         | –                 | 0.81 (0.65)         | 0.74 (0.58)          |

Note. BR = behavioral rehearsal; CSR = chart-stimulated recall.

*aReflects the number and percentage of participants endorsing this identity.*
were conducted at the first participating agency. The guide was subsequently adjusted to focus on the content and preserve the approximate one-hour time limit.

**Data Analysis**

*Quantitative Analyses of Questionnaire Data.* First, for each method, and consistent with Aim 1, we report on descriptive analyses for the four TPB variables: attitudes, self-efficacy, norms, and intention. Second, we conducted paired samples and independent samples t-tests to examine if supervisor or clinician attitudes, self-efficacy, norms, and intention toward BR and CSR differed. Third, and consistent with Aim 2, to ensure that these constructs explained variance in intention as expected, we conducted a linear regression using bootstrapped p-values (Austin & Tu, 2004) to test if attitudes, self-efficacy, norms, and condition predicted intention in the clinician sample. The sample size was too small for the supervisor sample to conduct a similar analysis.

*Qualitative Analysis of Belief Elicitation Interview Data.* In service of our second aim, we identified which beliefs were most commonly reported when asked each of the elicitation questions. Using a content analysis approach, coders summed the frequency for which specific beliefs were reported (Middlestadt et al., 2011; Middlestadt et al., 2013). For example, when analyzing self-efficacy beliefs, the coders tracked how often respondents reported that “time” would make it difficult to use BR and/or CSR routinely. As another example, when analyzing normative beliefs, the coders tracked how often supervisors were mentioned when asked who would approve or disapprove of them using BR and/or CSR routinely. Two individuals separately coded responses and demonstrated high intercoder reliability. Coders had an average of 96% agreement (range = 92%–98%). We report the beliefs most commonly endorsed (i.e., mentioned by at least 15% of participants).

In service of our third aim, we employed an integrated approach to organizing the belief elicitation interview data using the CFIR to identify contextual factors not captured by TPB-focused analyses. Four members of the research team reviewed five individual transcripts, identifying themes independently, and convened to compare preliminary codes. The final codebook (available upon request) contained 15 unique, neutral codes which were mapped onto the five CFIR domains (Intervention Characteristics, Outer Setting, Inner Setting, Individuals Involved, Process). The same codebook was used for supervisor and clinician interviews. Once the codebook was finalized, supervisor and clinician transcripts were coded by four coders. The coding team met weekly to review questions and reach consensus. Four transcripts were coded independently and reviewed as a group to achieve reliability (reliability rate ≥ 0.80); transcripts then were coded independently with a subset of 12 transcripts overlapping between coders to assess final reliability across codes (reliability rate ≥ 0.80).

**Results**

**TPB Quantitative Survey Findings**

See Table 2 for descriptive statistics for attitudes, self-efficacy, norms, and intention.

*Intentions.* There were no significant differences between supervisors’ intention to use BR or CSR. There were no differences in clinicians’ intention to use BR or CSR. For both types of participants, intentions were fairly high.
Table 2. t-tests testing attitudes, self-efficacy, norms, and intentions in supervisors and clinicians.

| Variable | BR n | M | SD | CSR n | M | SD | t(14) | p       |
|----------|------|---|----|-------|---|----|-------|---------|
| Attitudes | 15   | 6.46| 1.87| 34 | 6.60| 2.66| 0.512 | .618    |
| Self-efficacy | 15 | 4.17 | 0.65 | 4.23 | 0.68 | 3.77 | 0.294 | .004   |
| Norms | 15 | 3.30 | 0.77 | 3.77 | 0.78 | 2.168 | .057   |
| Intention | 14 | 5.07 | 1.49 | 5.43 | 0.65 | 0.751 | .468    |

Table 3. Paerson correlations between attitudes, self-efficacy, norms, and intention in clinicians.

| Variable | Attitudes | Self-efficacy | Norms | Intention |
|----------|-----------|--------------|-------|-----------|
| Attitudes | –         | 0.099        | 0.375* | 0.465*    |
| Self-efficacy | 0.099       | –           | 0.135 | 0.033     |
| Norms | 0.375* | 0.135        | –     | 0.429*    |
| Intention | 0.465* | 0.033        | 0.429* | –         |

Note. BR = behavioral rehearsal; CSR = chart-stimulated recall.

Table 4. Clinician attitudes, self-efficacy, norms, and condition predicting intention.

| Variable   | B     | SE   | B     | p     |
|------------|-------|------|-------|-------|
| Attitudes  | 0.196 | 0.064| 0.355*| .036  |
| Self-efficacy | –0.112 | 0.263| –0.046| .706  |
| Norms      | 0.563 | 0.218| 0.301*| .031  |
| Condition  | –0.062| 0.297| –0.023| .843  |

Note. R^2 = 0.294; B = unstandardized coefficient; B = standardized coefficient.
*p < .05.

Beliefs Underlying Attitudes

Commonly Reported Dislikes/Disadvantages

TPB-guided coding highlighted dislikes/disadvantages of implementing either fidelity measurement method that spanned various CFIR categories. In the intervention characteristics domain, supervisors and clinicians expressed concern that BR may not give an accurate portrayal of what happened in the session. For example, clinicians reported that their memory may worsen if BR occurs too long after the session, that role-plays may not adequately capture client or clinician behavior, and supervisors, as well as clinicians, were concerned that BR may not be able to provide enough context about what happened. Concern about memory lapse was also present for CSR, but supervisors and clinicians agreed that having session notes mitigated this concern. In the domain of Inner Setting factors, clinicians worried that their supervision time would be less helpful if these tools were used.

Qualitative Belief Elicitation Interview Results

Themes that emerged from TPB coding and CFIR coding overlapped, such that beliefs coded within attitude, self-efficacy, and normative constructs (Pathway B, Figure 1) were captured within the CFIR’s coding of contextual factors (Pathway A, Figure 1). Because CFIR coding sought to categorize the responses in a manner consistent with implementation science terminology, and because the TPB method prioritizes lists of factors, findings from CFIR coding elaborated rich contextual information informing the specific beliefs identified by TPB coding. Due to this overlap, CFIR findings are presented within the TPB constructs they underly (Table 5).
Table 5. Theory of planned behavior and consolidated framework for implementation research integration of results.

| CFIR domains | TPB constructs | Beliefs underlying attitudes | Normative referents | Self-efficacy beliefs |
|--------------|----------------|-------------------------------|--------------------|----------------------|
| Inner setting | Supervision time | Dislikes/disadvantages | Likes/advantages | Inner setting time may | There is already a shortage of |
| Either fidelity method | become less helpful than it is now.\(^a\) | | | time to complete | important tasks in |
| | | | n/a | supervision (e.g. | administrative tasks, clinical |
| | | | | skill building, etc.\(^a\)) | |
| | | | | For client crisis | Fidelity measurement may |
| | | | | sessions, | not be perceived as a good |
| | | | | fidelity measurement | use of time.\(^b\) |
| | | | | may make clinicians | Frequent disruptions (walk-in |
| | | | | consistent execution of | clients) may complicate |
| | | | | a method.\(^b\) | consistent execution of |
| | | | | Clinicians are already | a method.\(^b\) |
| | | | | overburdened by current | |
| | | | | demands and may not have | |
| | | | | the capacity to incorporate | |
| | | | | additional supervision | |
| | | | | activities.\(^a\) | |
| BR specific | n/a | n/a | n/a | n/a | Additional effort is required |
| CSR specific | n/a | n/a | n/a | n/a | to ensure that session |
| | | | | notes are prepared for | |
| | | | | supervisor to review in | |
| | | | | supervision.\(^b\) | |
| Intervention characteristics | Measure may not give an accurate portrayal of what transpired in session due to memory lapse, inability to communicate complexity of the session.\(^a\) | | Fidelity method may make clinicians more aware of when they use CBT in session\(^b\) | n/a | |
| Either fidelity method | Roleplaying with a supervisor can be awkward\(^b\) | | | Clinicians felt clients may disapprove if they knew BR was implemented.\(^b\) | n/a |
| | | | | Clinicians identified that other clinicians may dislike CSR due to tension, “interrogative” nature. (5XY004)\(^b\) | n/a |
| CSR specific | Perceived “interrogative” nature of CSR format may cause clinicians discomfort. (5XY004)\(^a\) | | CSR may make clinicians more mindful of why CBT is useful.\(^b\) | | |
| | | | | Clinicians identified that other clinicians may dislike CSR due to tension, “interrogative” nature. (5XY004)\(^b\) | n/a |
| | | | | Clinicians identified that other clinicians may dislike CSR due to tension, “interrogative” nature. (5XY004)\(^b\) | n/a |
| Individuals Involved | Measures may be overly structured, causing supervision to feel “rote” or “robotic” (4SN004, 4QU004)\(^a\) | | Measure may benefit supervision time by lending structure\(^a\) | Clinic leadership would approve of implementing a fidelity method as it would promote EBP uptake\(^a\) | Using a fidelity measurement may not be appropriate for |
| Either fidelity method | Assessment may detract from modeling and mentorship currently taking place in | | Incorporating a method may reinforce clinician knowledge about how to use CBT techniques (or help clinicians learn CBT techniques)\(^a\) | | all clinicians: Seasoned clinicians may benefit less than newer clinicians who are still familiarizing themselves with CBT. However, newer clinicians may benefit less|

(Continued)
First, using a fidelity check may apply pressure to change the behavior of clinicians in session with their clients. Supervisors and clinicians viewed this as an infringement on clinician autonomy, as evidenced in the following quote from a supervisor speaking about CSR, “So I think if I had a list of did you do this or that, it would take away from the clinician’s feeling like they had some autonomy to do what they felt was appropriate in the session at the time” (4WH004). Second, they felt that introducing these methods may create a threatening sense of constant evaluation or “looking for problems rather than...growth opportunities” (4QJ004), although some added that they valued role-play experiences for the learning opportunities they provide despite this discomfort. One supervisor pointed out that performance reviews currently depend upon professionalism and general good standing as an employee, and expressed concern that using CBT fidelity assessment may begin to factor into performance appraisals (which they viewed unfavorably). Clinicians also identified this tension as a concern, although discomfort around the “interrogative” (5XY004) nature of a method was only voiced regarding CSR, while anxiety in BR seemed to refer to awkwardness of role-playing with a supervisor.

**Commonly Reported Likes/Advantages**

Participants also identified the advantages of using the methods. In the CFIR domain of Individuals Involved, supervisors and clinicians hypothesized that incorporating

---

Table 5. (Continued)

| CFIR domains | TPB constructs |
|--------------|---------------|
| **Dislikes/disadvantages** | **Likes/advantages** | **Normative referents** | **Self-efficacy beliefs** |
| Supervision | Opportunity for rich feedback | Determined by supervisors and clinicians as they need supervision time to develop general clinical skills, learn proper documentation procedures, etc. |
| Pressure on clinicians to perform or a sense of constant critical evaluation, resulting in anxiety and detriment to supervisor rapport | | | |
| Perceived infringement on clinician autonomy in session | | | |
| **BR specific** | BR may not give an accurate portrayal of what transpired in session due to supervisor inability to adequately act as client | Value role play as a training tool despite discomfort | n/a | n/a |
| **CSR specific** | Memory lapse may lead to inaccuracy | Ability to use session note mitigates risk of memory lapse | n/a | n/a |
| **Outer Setting** | **Either fidelity method** | Fee-for-service clinicians may be unable to execute this kind of task given that they are not paid for their time outside of session and their schedules may be unwieldy |
| | n/a | n/a | n/a | n/a |
| **BR specific** | n/a | n/a | n/a | n/a |
| **CSR specific** | n/a | n/a | n/a | n/a |

Note. Response’s orientation within CFIR coding indicated by row, TPB orientation indicated by column. Qualitative findings indicate the relationship between CFIR-identified contextual factors and TPB constructs as represented by Pathway A (Figure 1). CBT = cognitive behavioral therapy; TPB = theory of planned behavior; BR = behavioral rehearsal; CSR = chart-stimulated recall; CFIR = Consolidated Framework for Implementation Research. 

Expressed by both supervisors and clinicians.

Expressed by clinicians.

Expressed by supervisors.
either method may reinforce clinician knowledge about how to use CBT techniques (or help clinicians learn to use CBT techniques) and provide an augmented opportunity for rich feedback in supervision. Specific to BR, clinicians identified the benefit of additional opportunities to reflect on what happened in session with their supervisor, while clinicians liked that CSR would make them more mindful of why CBT is useful. In the domain of Intervention Characteristics, clinicians liked that using these methods routinely would make them more aware of the CBT that they do use in the session.

Commonly Reported Self-Efficacy Beliefs

Supervisors and clinicians expressed concerns about their ability to execute either method for reasons that primarily mapped onto the CFIR domains of Inner Setting, Outer Setting, and Individuals Involved. Regarding Inner Setting factors, supervisors and clinicians reported that supervision time is a precious commodity, and this time is frequently dedicated to reviewing clinical concerns, other modalities of therapy, and administrative issues/agency procedures. Clinicians stated that they already lack time in supervision to discuss necessary topics, particularly for clients in crisis. Using a new measure that would take time from these priorities caused clinicians concern that they may “waste” their supervision time (5AM004), as this intervention may “get in [the] way of what [they] actually need from supervision,” (5AM004). Between the expectations around session time and documentation, clinicians often feel they do not have time to complete certain exercises that are currently mandated by the clinic, stating that “we already don’t do the mandated stuff … I think adding one more thing, I just don’t think anyone would do it well,” (5AM004). Specific to CSR, clinicians expressed concern related to the need for session notes to be prepared, available, and ready for supervisor review to execute the method. These tasks constituted “just one more thing we have to do” in an already busy schedule (5JQ006).

Relatedly, in the CFIR domain of Outer Setting, clinicians expressed concern about either method in terms of how they may impact fee-for-service clinicians. Because these clinicians are often not paid for time that is not spent with clients, every additional task on their plate constitutes unpaid work. Adding time to supervision that may not be perceived as contributing directly to their growth as a clinician (as time spent focusing on development of clinical skills or necessary administrative topics may) and for which they are not paid constituted an understandable burden. Additionally, fee-for-service clinicians often work at multiple sites, making the logistics of scheduling time to execute this kind of supervision task difficult.

In the CFIR domain of Individuals Involved, supervisors and clinicians felt that implementing these methods may not be equally appropriate for all clinicians. Some felt that more seasoned clinicians would benefit less from these methods than newer clinicians who are still familiarizing themselves with CBT, while others believed that it would be less appropriate to use these methods with a new clinician who should spend supervision time learning more basic elements of therapy (e.g., rapport-building, proper documentation). Supervisors and clinicians thought that using these methods would be more difficult or less appropriate for complex or highly traumatized clients, as they would not be able to give sufficient context to communicate a comprehensive idea of the session.

Commonly Reported Normative Referents

Supervisors and clinicians thought that clinic leadership (e.g., upper management, clinical leads, supervisors) would approve of implementing BR or CSR as it would promote CBT uptake in their organization, a perceived benefit. Clinicians expressed concern that their clients may disapprove if they knew that BR was being implemented, as they feared their clients may dislike the concept of someone acting as them during supervision. Clinicians believed that other clinicians may dislike using CSR in supervision, citing the same reasons expressed as dislikes surrounding this method (e.g., tension, perceived interrogative nature of CSR). As these comments pertained to the opinions or evaluations of others, all of these observations coded by TPB as normative referents were mapped onto the CFIR domain of Individuals Involved.

Discussion

We examined intention and determinants of intention to use two promising methods to index adherence to CBT in community mental health. We did so using approaches that include a causal theory (TPB) and a leading implementation science determinant framework (CFIR) with hopes of establishing a richer understanding of contextual factors that underlie beliefs. We hypothesized that stakeholders would have a stronger intention and perceive fewer barriers to using CSR compared to BR. Quantitative measurement suggests that intentions to use both methods were moderately strong and not significantly different from each other or across supervisors and clinicians. Attitudes and norms were related to the clinician intention to use the methods, suggesting potential targets for future strategies to strengthen intentions toward using these methods. Qualitative data highlights the specific beliefs and the contextual factors informing them that should be influenced to change attitudes and norms. These insights suggest that there are many multi-level factors influencing individual attitudes, self-efficacy, and norms that must be addressed prior to implementation of any method to measure components of fidelity, including adherence (see Pathway A, Figure 1).

Given that clinician intentions to use BR and/or CSR were related to attitudes and norms (Pathway B, Figure 1), effective and tailored implementation strategies to increase
Clinicians’ use of BR and/or CSR could target these specific beliefs. For example, an implementation strategy targeting attitudes could include messages to change beliefs about perceived advantages and disadvantages to using such methods. The field of social psychology has developed messages that can change attitudes, which thus strengthens intention and changes behavior (Fishbein & Ajzen, 2010). Strategies can target norms with communication-based interventions to leverage message effects (Fishbein & Yzer, 2003; Hornik & Woolf, 1999). For example, if norms explain the bulk of variance in intention, messages that change one’s perception of the norms have been effective in various contexts (Cialdini et al., 2006).

Understanding the contextual factors that influence implementation is a core activity of implementation science. First, although clinicians and supervisors implicitly described how the outer setting, or broader sociopolitical context, impacted their ability to use these fidelity measurement methods, they did not explicitly or routinely mention these factors in the belief elicitation interview with the exception of financial burden imposed by the payment structure of fee-for-service clinicians. Given the overburdened nature of community mental health, it is critical to note that these fidelity methods will likely not be implemented unless they are financially incentivized, mandated by payers, and/or incorporated into standard workflow. Had we included organizational directors or other higher-level policy makers (county or state mental health leaders), these themes might have emerged more consistently.

Second, clinicians and supervisors raised concerns about the methods themselves, with a particular focus on whether either could accurately portray what happened in session and the potential loss of focus in supervision that may accompany implementation of a tool focused on CBT (in lieu of other modalities or approaches). Likely, this concern relates to attitudinal perceptions of the acceptability and the feasibility of these methods (Proctor et al., 2011). One implementation strategy to address these perceptions includes training to enhance supervisor and clinician understanding of how these methods can and should be used (Beidas et al., 2012). Another strategy would be to partner with community stakeholders to redesign approaches to be more amenable to the realities of community mental health and ensure their acceptability and feasibility (Lyon & Bruns, 2019). Although we endeavored to do so given our longstanding experience in community mental health, co-creation of the measures may have resulted in more acceptable approaches (Metz, 2015). To address these concerns, implementation efforts may need to target buy-in around the accuracy of BR and CSR and opportunities created when using these methods for growing clinicians’ skills in CBT. It may also be helpful to describe client benefits of delivering CBT with fidelity. One strategy that may broadly address these targets could include education of key clinic stakeholders around the accuracy of BR and CSR and the known benefit of CBT for youth.

Third, concerns were raised about the characteristics of the inner setting, or the clinics in which community mental health is delivered, particularly related to time and the chaotic nature of community mental health delivery (Stewart et al., 2021). Much of these inner setting challenges are likely due to systemic outer setting factors such as poor reimbursement. Community mental health workers are often overburdened and undercompensated, resulting in a highly stressful working environment (Beidas et al., 2016a). While incorporating BR or CSR into regular practice may increase the quality of treatment, addition of new demands could increase stress surrounding burden and potentially lead to burnout. This risk may be particularly poignant for fee-for-service model employees as they typically are not compensated for such activities that fall outside of clinical work (Beidas et al., 2016b).

Fourth, with regard to individuals involved, a number of factors related to clinician autonomy and characteristics came up, with some interesting suggestions that these types of methods are not appropriate for more seasoned clinicians or complex sessions. These themes are corroborated by the literature on implementation of evidence-based practices (McHugh & Barlow, 2010), and specifically CBT. Participants expressed that these methods do not adequately capture the complexity of what is needed in community mental health. While it is understandable that one evidence-based practice like CBT may not be sufficient to meet the complex needs of community mental health care delivery, we argue that it is imperative from an ethics perspective to ensure that we are delivering the highest quality care in these settings, which currently includes CBT approaches.

In this work, we endeavored to bring together an established causal theory from social psychology, the TPB, and an established framework from implementation science, the CFIR, consistent with recent calls to grow our understanding of how theories and frameworks operate rather than reifying existing approaches (Kislov et al., 2019). Interestingly, there was strong convergence between the two approaches, with the TPB guiding our methods for eliciting beliefs about using each method, and the CFIR offering an organizing structure that allowed for more meaningful insights and richer contextual understanding to inform future implementation strategy selection (Waltz et al., 2019). Aligning terminology and approaches was not without challenges, most notable of which was the orientation of CFIR findings within TPB results to avoid redundancy. Given the conceptual convergence in coding results, decisions about which context-rich CFIR findings presented a novel, necessary context and which findings were sufficiently represented in reported TPB results presented a particularly complex challenge that could only be navigated through multiple iterations of integration and critical feedback from investigators with varying expertise. Bringing together these approaches required collaboration between scientists from different disciplines.
(Falk–Krzesinski et al., 2010), and allows for a merging of approaches that advance our understanding. Limitations of the current work include the following. First, with interviews and surveys, there is potential for social desirability bias. This appears to have been limited given that clinicians reported both negative and positive beliefs about each fidelity measurement method. Second, supervisors did not have direct experience with the methods that they rated, which may have influenced their response style. Third, this work is hypothetical given that these methods are not currently mandated by systems. Finally, as we did not examine actual supervisor and clinician use of these methods, we can only speak directly to factors informing intention (Pathways A and B, Figure 1). Future research should examine the potential moderation effect of contextual factors on the intention-behavior gap (Pathway C, Figure 1).

In conclusion, if these methods are to achieve their promise, they will need to be implemented in community mental health settings. This work elucidates key targets and implementation strategies that will need to be deployed to take these methods to scale in community mental health, and illustrates a process that can be used more broadly in implementation science merging the TPB and CFIR approaches.

Acknowledgments

The authors are grateful for the support that the Department of Behavioral Health and Intellectual Disability Services has provided us to conduct this work within their system, for the Evidence Based Practice and Innovation (EPIC) group, and for the partnership provided to us by participating agencies, clinicians, and clients. The authors thank Kelly Zentgraf, MS, Courtney Gregor, MS, and Mary Phan, BA, for their roles as research coordinators on this project over the study period, as well as Steven Marcus, PhD, and Eliza Maceal, M.Sc. for their assistance with statistical analyses.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs Sonja Schoenwald and Rina Beidas both serve as co-authors on this paper, and have editorial affiliations with the Journal. Dr. Schoenwald is Co-Founding Editor in Chief of Implementation Research and Practice. Dr. Beidas was previously an Associate Editor; all decisions on this manuscript were made by another editor. Dr. Beidas is principal at Implementation Science & Practice, LLC. She receives royalties from Oxford University Press, consulting fees from United Behavioral Health and OptumLabs, and serves on the advisory boards for Optum Behavioral Health, AIM Youth Mental Health Foundation, and the Klingenstein Third Generation Foundation outside of the submitted work.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of Mental Health, (grant number R01 MH108551).

ORCID iDs

Carlin Hoffacker https://orcid.org/0000-0001-8939-9625
Sonja K. Schoenwald https://orcid.org/0000-0002-0560-8440
Shannon Dorsey https://orcid.org/0000-0001-6240-0294

Trial Registration

This study was registered with ClinicalTrials.gov (Identifier NCT02820623).

References

Becker-Haimes, E. M., Marcus, S. C., Klein, M., Schoenwald, S. K., Dorsey, S., Mandell, D.S., McLeod, B. D., Fugo, P. B., & Beidas, R. S. (2020, December 15-17). A randomized trial to identify accurate and cost-effective fidelity measurement methods for cognitive-behavioral therapy: Project FACTS preliminary findings. [Conference Presentation], 13th Annual Conference on the Science of Dissemination and Implementation in Health.

Becker-Haimes, E. M., Marcus, S. C., Klein, M. R., Schoenwald, S. K., Fugo, P. B., McLeod, B. D., Dorsey, S., Williams, N. J., Mandell, D. S., & Beidas, R. S. (2022). A randomized trial to identify accurate measurement methods for adherence to cognitive-behavioral therapy. Behavior Therapy, Advance online publication. https://doi.org/10.1016/j.beth.2022.06.001

Austin, P. C., & Tu, J. V. (2004). Bootstrap methods for developing predictive models. The American Statistician, 58(2), 131–137. https://doi.org/10.1198/0003130043277

Beidas, R. S., Stewart, R. E., Adams, D. R., Fernandez, T., Lustbader, S., Powell, B. J., Aarons, G. A., Hoagwood, K. E., Evans, A. C., Hurford, M. O., Rubin, R., Hadley, T., Mandell, D. S., & Barg, F. K. (2016a). A multi-level examination of stakeholder perspectives of implementation of evidence-based practices in a large urban publicly-funded mental health system. Administration And Policy In Mental Health, 43(6), 893–908. https://doi.org/10.1007/s10488-015-0705-2

Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don’t tell me: behavioral rehearsal as a training and analogue fidelity tool. Cognitive and Behavioral Practice, 21(1), 1–11. https://doi.org/10.1016/j.cbpra.2013.04.002

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

Beidas, R. S., Stewart, R. E., Benjamin Wolk, C., Adams, D. R., Marcus, S. C., Evans, A. C., Jackson, K., Neimark, G., Hurford, M. O., Erney, J., Rubin, R., Hadley, T. R., Barg, F. K., & Mandell, D. S. (2016b). Independent contractors in public mental health clinics: implications for use of evidence-based practices. Psychiatric Services (Washington, DC), 67(7), 710–717. https://doi.org/10.1176/appi.ps.201500234

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. Social Influence, 1(1), 3–15. https://doi.org/10.1080/15534510500181459
Middlestadt, S. E., Stevenson, L. D., Hung, C. L., Roditis, M. L., Fly, A. D., & Sheats, J. L. (2011). Beliefs underlying the decision to eat breakfast: the role of theory-based behavioral analysis in the development of policy, communication and educational interventions for healthy eating. Food Studies, 1(4), 45–54. PMID: 24089658

Perepletchikova, F., Treat, T.A., & Kazdin, A.E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. Journal of consulting and clinical psychology, 75(6), 829–841. https://doi.org/10.1037/0022-006X.75.6.829

Proctor, E. K., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G. A., Benger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Administration and Policy in Mental Health and Mental Health Services Research, 38(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Salvatori, P., Simonavicius, N., Moore, J., Rimmer, G., & Patterson, M. (2008). Meeting the challenge of assessing clinical competence of occupational therapists within a program management environment. The Canadian Journal of Occupational Therapy, 75(1), 51–560. https://doi.org/10.1177/000841740807500109

Sheeran, P. (2002). Intention—behavior relations: A conceptual and empirical review. European Review of Social Psychology, 12(1), 1–36. https://doi.org/10.1080/14792772143000003

Stewart, R.E., Mandell, D.S., & Beidas, R.S. (2021). Lessons from maslow: prioritizing funding to improve the quality of community mental health and substance use services. Psychiatric Services, 72(10), 1219–1221. https://doi.org/10.1176/appi.ps.20200209.

Sutton, S., French, D.P., Hennings, S.J., Mitchell, J., Wareham, N.J., Griffin, S., Hardeman, W., & Kinmonth, A.L. (2003). Eliciting salient beliefs in research on the theory of planned behaviour: the effect of question wording. Current Psychology, 22(3), 234–251. https://doi.org/10.1007/s12144-003-1019-1

Waltz, T.J., Powell, B.J., Fernández, M.E., Abadie, B., & Damschroder, L.J. (2019). Choosing implementation strategies to address contextual barriers: diversity in recommendations and future directions. Implementation Science, 14(1), 42. https://doi.org/10.1186/s13012-019-0892-4

Webb, T.L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin, 132(2), 249. https://doi.org/10.1037/0033-2909.132.2.249

Wolk, C.B., Becker-Haimes, E.M., Fishman, J., Affrunti, N.W., Mandell, D.S., & Creed, T.A. (2019). Variability in clinician intentions to implement specific cognitive-behavioral therapy components. BMC Psychiatry, 19(1), 406. https://doi.org/10.1186/s12888-019-2394-y