



# HHS Public Access

Author manuscript

*Adm Policy Ment Health*. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

*Adm Policy Ment Health*. 2022 March ; 49(2): 237–254. doi:10.1007/s10488-021-01160-4.

## Self-Coding of Fidelity as a Potential Active Ingredient of Consultation to Improve Clinicians' Fidelity

EB Caron, Ph.D.<sup>a</sup> [Assistant Professor], Mary Dozier, Ph.D.<sup>b</sup> [Amy E. du Pont Chair of Child Development]

<sup>a</sup>Department of Psychological Science, Fitchburg State University, 160 Pearl St., Fitchburg, MA 01420

<sup>b</sup>Department of Psychological and Brain Sciences, University of Delaware, 108 Wolf Hall, Newark, DE 19716

### Abstract

**Purpose:** A key goal for implementation science is the identification of evidence-based consultation protocols and the active ingredients within these protocols that drive clinician behavior change. The current study examined clinicians' self-coding of fidelity as a potential active ingredient of consultation for the Attachment and Biobehavioral Catch-up (ABC) intervention. It also examined two other potential predictors of clinician fidelity in response to consultation: dosage of consultation and working alliance.

**Method:** Twenty-nine clinicians (97% female, 62% White,  $M$  age = 34 years) participated in a year of weekly fidelity-focused ABC consultation sessions, for which clinicians self-coded fidelity and received consultant feedback on both their coding and their fidelity. Data from the ABC fidelity measure were available for 1067 sessions coded by consultants, and clinicians' self-coding accuracy was calculated from 1044 sessions coded by both clinicians and consultants. Alliance was measured with the Working Alliance Inventory - Trainee and Supervisor Versions. The study was observational, and fidelity and self-coding accuracy were modeled across time using hierarchical linear modeling.

**Results:** Clinicians' ABC fidelity, as well as their self-coding accuracy, increased over the course of consultation. Clinicians' self-coding accuracy predicted their initial fidelity and growth in fidelity. Working alliance was also linked to fidelity and self-coding accuracy.

---

Correspondence concerning this article should be addressed to EB Caron, Department of Psychological Science, Fitchburg State University, 160 Pearl St., Fitchburg, MA 01420, USA. [ecaron3@fitchburgstate.edu](mailto:ecaron3@fitchburgstate.edu); phone: 1-978-665-4144.

#### Data

In order to maximize sample size in the current study, the dataset used in Caron & Dozier (2019) was included in the current study, representing 6 of the 29 clinicians. Preliminary data from the current study were published as a dissertation (Caron, 2017), but have not previously been published in a peer-reviewed journal.

#### Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

#### Ethical Approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Institutional Review Board of the University of Delaware.

#### Informed Consent

Informed consent was obtained from all clinicians and consultants included in the study.

**Conclusions:** These results suggest that clinician self-coding should be further examined as an active ingredient of consultation. The study has important implications for the design of consultation procedures and fidelity assessments.

### Keywords

Consultation; fidelity; active ingredients; self-coding; implementation

---

About one in every five Americans suffers from mental illness, and nearly half of these individuals do not receive treatment (Child and Adolescent Health Measurement Initiative, 2019; Institute of Medicine, 2015). Many interventions have shown efficacy in treating mental illness in randomized controlled trials (Weisz et al., 2017). Yet when these evidence-based treatments are implemented in community settings, they frequently perform more poorly than expected, a discrepancy often attributed to low treatment fidelity (Cox et al., 2020; Hulleman & Cordray, 2009; Weisz et al., 2013), although associations between fidelity and outcomes have been mixed (Collyer et al., 2020). Implementation science studies the processes that can successfully integrate and improve outcomes of evidence-based treatments in community settings (Proctor et al., 2009).

Implementation frameworks offer a systematic model to conceptualize implementation processes, and can enhance intervention effectiveness by identifying key targets for behavior change (Tabak et al., 2012). The implementation framework developed by the National Implementation Research Network (NIRN) identifies two levels of processes that promote behavior change (Fixsen et al., 2005). First, core intervention components are the active ingredients of treatment that lead to client behavior change, or “any therapeutic skill, process or component with a demonstrated relationship to outcome or mediator of outcome” (Magill et al., 2015). Examples of core intervention components include assignment of client homework (Kazantzis et al., 2016) and in vivo feedback about parenting skills (Barnett et al., 2014; Caron et al., 2018). Second, core implementation components are the active ingredients that lead to clinician and agency behavior change, such as training and consultation. When core intervention components are understood, core implementation components can focus on the aspects of an intervention that matter most for outcomes. For example, identification of core intervention components can refine fidelity measurement, training, and consultation procedures to focus on intervention processes that are specifically linked to client change. A key aspect of the NIRN implementation framework is a feedback loop, in which clinicians’ fidelity is measured to provide feedback to consultants and other stakeholders about implementation progress. Rooted in this framework, the current study examines potential active ingredients of a core implementation component, fidelity-focused consultation, and its built-in feedback loops to consultants and clinicians.

Consultation is externally-provided support that begins after training and continues while clinicians are implementing an intervention (McLeod et al., 2018). Although consultation generally promotes clinician behavior change (Edmunds, Beidas et al., 2013), some studies have failed to find that consultation improved clinician fidelity (e.g., Moyers et al., 2008). Other studies have found benefits for certain types of consultation but not others (e.g., Funderburk et al., 2015) and links between specific consulting activities and

clinician fidelity (e.g., Schoenwald et al., 2004). In particular, active learning processes such as modeling, role play, and feedback have been linked with improved intervention implementation (Bearman et al., 2013; Bearman et al., 2017; Edmunds, Kendall et al., 2013). Thus, consultation on the whole is not necessarily an active ingredient of implementation; rather, specific types of consultation and strategies within consultation have been linked to clinician behavior change. Developing an evidence base for active ingredients and models of consultation is a critical goal for implementation science, as this work could promote effective translation of evidence-based treatments into community settings (McLeod et al., 2018).

One gold standard model of consultation involves providing feedback to clinicians on their fidelity as coded from recorded sessions (Dunn et al., 2016). This type of consultation has been found effective in promoting clinician fidelity in various interventions, including motivational interviewing (Martino et al., 2016), cognitive-behavioral therapy (Weck et al., 2017), and Coping Power (Eiraldi et al., 2018). Consultation procedures that involve both consultants and clinicians themselves coding fidelity have been rare in the field, but are theorized to enhance clinician fidelity by sensitizing clinicians to certain aspects of fidelity and opportunities to use intervention strategies with clients (Isenhardt et al., 2014). Isenhardt et al. (2014) found that two clinicians showed increasing fidelity to motivational interviewing over time while they were engaged in weekly group supervision that included coding session segments as a group.

Caron and Dozier (2019) examined the effects of “fidelity-focused consultation,” in which both consultants and clinicians completed weekly fidelity coding of sessions, among 7 community-based clinicians implementing Attachment and Biobehavioral Catch-up (ABC), an evidence-based parent coaching intervention for infants. In this multiple baseline design, clinicians first received weekly group-based consultation as usual, and then began receiving weekly supplemental individual fidelity-focused consultation provided by undergraduate student fidelity coders, with ongoing consultation as usual. The clinicians showed increasing fidelity after the onset of fidelity-focused consultation, but not during the consultation as usual period, suggesting that fidelity-focused consultation improved clinicians’ fidelity. It is important to extend this work and examine possible active ingredients of consultation that are linked to improvements in fidelity, which would allow consultation strategies to be extracted and adapted for use with other evidence-based treatments.

One possible active ingredient is clinician self-coding. Self-coding of fidelity is a form of reflection or reflective practice (Caron et al., 2021), which is theorized to be a key driver of learning for both trainees and experienced clinicians (Bennett-Levy et al., 2009). Caron and Dozier (2019) argued that when clinicians engage in self-coding of fidelity, they may be more receptive to consultants’ fidelity feedback, due to increased perceived source credibility and feedback validity. Self-coding of fidelity may also provide clinicians with the skills to self-monitor following cessation of consultation, and may promote sustainment of fidelity; in fact, Caron and Dozier (2019) found that clinicians’ fidelity remained stable for two years after fidelity-focused consultation ended.

However, much prior literature has suggested that clinicians are inaccurate, positively biased self-raters, while noting that differences in training and rating methods may contribute to clinicians' rating discrepancies (e.g., Carroll et al., 1998; Martino et al., 2009). Specifically, clinicians typically have completed retrospective self-reports, often with minimal training or feedback, whereas independent observers were thoroughly trained and based ratings on session recordings (Carroll et al., 1998; Martino et al., 2009). Although much of the work on fidelity self-rating has theorized that training in fidelity rating should increase clinicians' reliability with observers (Carroll et al., 1998; Hogue et al., 2015; Peavy et al., 2014), there is currently little research that has investigated this topic. In the field of cognitive behavior therapy, Loades and Myles (2016) found that 13 trainees trended toward demonstrating lower self-rating discrepancies with experts over time during a 30-week training course, and Beale et al. (2020) also found that 150 trainees demonstrated improved agreement over time during a yearlong training program. In addition, Hogue et al. (2021) found that 48 community therapists showed improvement in rating family therapy techniques over 32 weeks of online training in which they rated video vignettes, rather than their own session videos. Although the existing literature suggests that rater training improves interrater reliability (e.g., Reichelt et al., 2003), additional work is needed to demonstrate that clinicians' self-rating ability can improve with training. Given that clinician-rated fidelity measures offer a pragmatic solution to reduce the costs and burden of fidelity measurement in community settings (Brookman-Frazer et al., 2021; Serfaty et al., 2020), this line of research could justify investment of resources in training clinicians to rate fidelity.

In addition, research should explore links between clinicians' self-rating ability and their observed fidelity (Hogue et al., 2021). Because self-rating is a form of reflective practice, self-rating ability may promote in-session practice change. In support of this idea, some previous work suggests that relative to less competent clinicians, more competent clinicians tend to be more accurate self-raters, at least at the group level (i.e., comparing groups of more and less competent clinicians; Beale et al., 2020; Brosan et al., 2008; Caron et al., 2020). In contrast, other work suggests that trainees with greater competence may be more self-critical and likely to underestimate performance than trainees with less competence (McManus et al., 2012), and associations between reflective practice ability and self-rating accuracy have been mixed (Hitzeman et al., 2020; Loades & Myles, 2016). Further exploration of these associations within individuals, and examination of the differential associations of self-rating ability with baseline fidelity and growth in fidelity over time, could provide greater support for self-rating as a potential active ingredient of consultation linked to change in fidelity.

When examining clinician self-rating as an active ingredient of consultation leading to change in fidelity, it is important to include potential alternative active ingredients, such as working alliance and dosage of consultation. Working alliance is theorized to enhance implementation outcomes like fidelity by enhancing clinician motivation to engage in learning (McLeod et al., 2018), but links between clinician-consultant working alliance and implementation outcomes have been mixed (Schoenwald et al., 2004; Wehby et al., 2012), and more research is needed, as processes related to working alliance with consultants likely differ from those with clients. Prior work has also found links between dosage of consultation and fidelity (e.g., Beidas et al., 2012; Schwalbe et al., 2014), suggesting that

the amount of time spent in consultation is important. Importantly, both working alliance and consultation dosage are possible alternative explanations for the multiple baseline results of Caron and Dozier (2019); specifically, the observed increase in fidelity following the addition of fidelity-focused consultation could be attributed to an additional supportive relationship with a consultant or receiving additional consultation time. Evidence for active ingredients of consultation, including self-coding, working alliance, and consultation dosage, could allow design of consultation procedures that focus on components with demonstrated links to fidelity, which may increase the effectiveness of consultation in changing clinician behavior and in turn improve client outcomes.

In the current study, we examined processes related to clinicians' growth in fidelity to Attachment and Biobehavioral Catch-up. We hypothesized that clinicians' self-coding accuracy would improve over the course of fidelity-focused consultation, and explored whether early self-coding accuracy was associated with initial fidelity or growth in fidelity over time. We also examined two other aspects of fidelity-focused consultation as predictors of growth in clinicians' fidelity to ABC and self-coding accuracy: clinician-consultant working alliance and consultation dosage.

## Method

### Participants

**Clinicians**—Participants included 29 clinicians from 20 agencies in 6 US states who were invited to participate in the current study while receiving ABC consultation, through a verbal invitation and explanation of the study during ABC group clinical consultation. Clinicians were primarily employed at private non-profit agencies with contracts to provide services for children in the state child welfare system, but others were employed directly by state or county Child Protective, Welfare, or Health and Human Services departments. Clinicians within the same state were typically part of the same training cohort, as several organizations had initiated ABC implementation efforts in multiple agencies around the state. Clinicians who were already involved in ABC training and consultation were approached about participating in the study, which was observational and did not affect their ongoing training, regardless of their decision of whether to participate. Four additional clinicians were approached but declined to participate; one additional clinician consented but was not included in analyses because of insufficient data due to leaving the agency early in training. The decision of whether to participate in the study was individual, and agencies were not involved in this decision. Full demographic data for one clinician were not available because she started but did not complete the questionnaires.

All clinicians but one (97%) were female. Most ( $n = 20$ , 69%) clinicians had Master's degrees; 6 (21%) had Bachelor's degrees, 1 (3%) had a Ph.D., and 1 (3%) had completed high school or less. Most ( $n = 18$ , 62%) clinicians were White, 6 (21%) were Black, 1 (3%) was Asian American, and 3 (10%) were more than one race. One White clinician and one Black clinician (7%) were Hispanic. On average, clinicians were 34 years old ( $SD = 7$ , range: 23 – 48). They reported having worked an average of 4.7 years ( $SD = 5.5$ ) in their current jobs, and 8.9 years ( $SD = 6.5$ ) in the field.

**Fidelity Coding Consultants**—Twenty-two fidelity coding consultants consented to participate and contributed fidelity coding and self-report measures of clinician-consultant alliance in the current study. All consultants but one (95%) were female. Most consultants were White/non-Hispanic ( $n = 14$ , 64%), 6 (27%) were White/Hispanic, and 1 (5%) was Black. Average age was 22 years old ( $SD = 1.6$ , range: 20-26). The majority ( $n = 14$ , 64%) were undergraduate students, and 8 (36%) were part- or full-time staff. Of the staff members, 6 had a Bachelor's degree, 1 had a Master's, and 1 had a Ph.D.

## Procedure

**Attachment and Biobehavioral Catch-up Intervention (ABC)**—ABC is a 10-session, home-based parenting intervention for infants who have experienced early adversity, such as abuse and neglect. ABC aims to promote children's developmental outcomes by enhancing parental sensitivity both when children are distressed (i.e., “nurturing”) and when they are not distressed (i.e., “following the lead”), increasing parents' joy and delight in their children, and reducing parental behavior that could be frightening (e.g., threatening, yelling). Compared with children assigned to a control intervention, ABC has been found to promote children's organized and secure attachment (Bernard et al., 2012), normalized patterns of the stress hormone cortisol (Bernard et al., 2015), and executive functioning (Lind et al., 2017). Benefits for children have been found several years after the intervention has ended (e.g., Garnett et al., 2020; Lind et al., 2017). Further, changes in parenting have mediated the effects of the intervention on several child outcomes, including vocabulary (Raby et al., 2019), cortisol production (Garnett et al., 2020), and compliance (Lind et al., 2020).

ABC is delivered in families' homes by clinicians called “parent coaches.” Parent coaches teach parents about targeted parenting behaviors in several ways including manual-guided conversation, video models, and video feedback. In terms of manualized content, nurturing care is targeted in sessions 1-2, following the lead is the focus of sessions 3-4, and intrusive and frightening behavior is highlighted in sessions 5-6, whereas sessions 7-10 are individualized to the parent's strengths and challenges (detailed information on the manualized intervention content can be found in Dozier et al., 2017). However, the core intervention component of ABC is in-the-moment commenting, which clinicians use to coach parenting behavior throughout all other intervention activities (Caron et al., 2018). For example, a clinician might comment, “He bumped his head and you picked him up and snuggled him right away – that's an example of nurturance.” Clinicians video record sessions for the purposes of video feedback to parents and for use in consultation.

**Clinician Training and Consultation**—Clinicians were trained in a two- to three-day training workshop by Ph.D.-level clinical trainers who were reliable on the fidelity coding system but who were not fidelity coding consultants. Training on the fidelity coding system was integrated into the workshop, and represented about 25% of the total training time. Clinicians were provided with an ABC manual and a fidelity coding manual. After training, they began consultation, which included both group clinical consultation for one hour weekly and individual fidelity-focused consultation for a half hour weekly. Both types of consultation were conducted remotely over video conferencing software that allowed review of session video. In group clinical consultation, clinicians met with Ph.D.-level consultants



who were ABC model experts and used session video review to support delivery of manual content and case-specific dynamics.

In fidelity-focused consultation – the focus of the current paper – clinicians met with fidelity coding consultants who assigned randomly-selected 5-minute video clips for fidelity coding from 1 to 3 recent session videos. After clinicians sent consultants their self-coding, they received consultants' independently rated coding in return, prior to consultation. In consultation, fidelity-focused consultants provided clinicians with feedback on both their coding and their fidelity. Consultants' coding feedback was designed to improve clinicians' understanding and use of the fidelity coding measure, described below, including both intervention-targeted parent behaviors and clinician responses to those behaviors (i.e., in-the-moment comments). For example, a consultant might provide feedback about an incorrectly-coded parent behavior by saying, "The next behavior, you coded as following the lead, but it's actually nurturance. Even though the child didn't cry, when he fell down, the mom said, 'Uhoh, you fell,' which shows the child that the mom noticed and is available to provide comfort if he needs it." Consultants' fidelity feedback was designed to move clinicians progressively toward meeting ABC certification criteria, by helping clinicians to improve their comment frequency and quality. For example, a consultant might give feedback on comment quality by saying, "You had one off-target comment in this clip, where you mislabeled the behavior target and said, "Great following the lead," in response to a behavior that I had coded as nurturance. Let's look at the video so I can show you why I see that behavior as nurturance." In addition to feedback, fidelity coding consultants were encouraged to use other active learning techniques, including modeling and engaging clinicians in commenting role-play.

For 6 of the clinicians who were part of a multiple baseline study (Caron & Dozier, 2019), fidelity-focused consultation began several months after they started implementing ABC and receiving group clinical consultation. The rest of the clinicians began both fidelity-focused and group clinical consultation soon after the training workshop and often before they began implementing ABC. The two groups were compared on all predictor variables, and the only significant difference that emerged related to the difference in when the groups began fidelity-focused consultation; specifically, the multiple baseline group received significantly fewer fidelity-focused consultation sessions prior to their first session of ABC ( $M = 0.0$ ;  $SD = 0.0$ ) than the rest of the sample ( $M = 3.0$ ,  $SD = 2.5$ ),  $t(22) = 5.93$ ,  $p < .001$ . However, when the groups were compared in terms of fidelity outcomes, the clinicians who participated in the multiple baseline had lower fidelity than those who did not, likely because they were trained earlier than the others and training procedures related to fidelity were refined over time. For this reason, study group was included as a covariate in linear growth models of fidelity outcomes.

Consultation was intended to occur weekly for one year, and during this time, clinicians received an average of 32.8 ( $SD = 8.0$ ) sessions of fidelity-focused consultation. Switches in clinician-consultant pairings occurred when fidelity coding consultants left the lab temporarily (e.g., summer break) or permanently (e.g., graduating, leaving job). Excluding brief periods of substitute consulting (i.e., 4 sessions or less), most clinicians ( $n = 16$ , 55%) received consultation from a single fidelity coding consultant during their training period.

However, 11 (38%) had two fidelity coding consultants during their training, and 2 (7%) had three fidelity coding consultants.

**Fidelity Consultant Training**—Fidelity coding consultants were trained to reliability in the coding system within a semester (1.5-hour weekly group meetings, plus individual coding practice for a total of 6-9 hours/week for 10-12 weeks). After completing a 5-video practice reliability set, coders met with the trainer for individual feedback, and then completed a reliability set of 10 videos. Coders were considered reliable if they achieved >70% item-level match with the master coder, averaged across the 10 videos, on the reliability set. Fidelity coding consultants were considered experts in ABC fidelity only, and not general ABC procedures. After achieving reliability in fidelity coding, consultants were provided with a fidelity-focused consultation manual and consultation training, and began observing others' sessions, prior to beginning consultation themselves. After beginning consultation, they participated in weekly group supervision meetings to review fidelity coding and identify areas of focus for consultation.

**Data Collection**—Clinicians' and consultants' fidelity coding spreadsheets were attached to weekly emails written by fidelity coding consultants to clinical ABC consultants and agency supervisors, and were archived and used in the current study. The study protocol was approved by the University of Delaware Institutional Review Board (IRB), and clinicians and fidelity coding consultants provided informed consent for use of archived materials and for completing online questionnaires about working alliance and demographics. Client consent was not required by the IRB; client consent for intervention, including consent to be videorecorded and for recordings to be used in consultation, was managed by agencies, and confidential client information, other than videorecorded images and voices, was not shared with ABC consultants. Data were collected between June 2012 and May 2017. Clinicians received a \$20 gift card in compensation for their questionnaire completion.

## Measures

**ABC Fidelity**—The ABC fidelity measure is designed to assess ABC adherence and competence through the frequency and quality of a core component of ABC, in-the-moment commenting (Caron et al., 2018). It involves coding both relevant parent behaviors and aspects of the clinician's response to them. Parent behaviors that are coded include: following the lead, not following the lead, nurturing, not nurturing, delight, and frightening. Each instance of a parent behavior signals an opportunity for the clinician to respond with an in-the-moment comment; if the clinician fails to comment, the behavior is coded as a missed opportunity. Behaviors are coded at the micro-level, sequentially, so repeated instances of the same parent behavior are each coded as separate opportunities to comment. If the clinician does comment, the comment is coded as on- or off-target; on-target comments are appropriate and accurate based on the parent behavior, whereas off-target comments may provide incorrect information, be poorly timed, or confuse the parent. In addition, the number of components of information included in the comment are coded from 0 to 3; components include (1) describing the parent's behavior (e.g., "He tapped the blocks together and you said, 'Tap tap.'"), (2) labeling the behavior target ("That's following his lead."), and (3) describing a long-term outcome the behavior can have on the child ("That'll



help him learn to regulate himself.”). Comments can include 0 components and still be on-target, for example, if they are not specific enough about the parent’s behavior (e.g., “Beautiful” or “Great responding to her.”).

Fidelity is coded from 5-minute video clips of ABC sessions using an Excel spreadsheet that automatically calculates summary statistics about the parent’s behavior and the clinician’s commenting. The current study used the total counts of each parent behavior and the four commenting variables involved in ABC certification criteria: the frequency of on-target comments (i.e., number of on-target comments per minute), the percentage of missed opportunities, the percentage of on-target comments, and the average number of components. All variables were used to assess clinicians’ self-coding accuracy. In addition, the frequency and percentage of on-target comments were chosen as fidelity outcomes, as both have predicted parent behavior change at both the clinician- and case-level in prior work using 5-minute clips (whereas percentage of missed opportunities and average number of components have predicted outcomes at the clinician-level, only), and appear reliable and valid measures of fidelity (Caron et al., 2018). Consultants’ coding was considered a proxy for the gold standard of “truth” and was used to assess fidelity outcomes. Data from 1067 sessions were available; on average, each clinician had data from 37 sessions ( $SD = 15$ , range: 12 – 76). Sixty-nine videos (6%) were double-coded by other fidelity coding consultants. The proportion of double-coded videos was limited by the time and resource constraints of implementation, and videos were assigned by convenience when consultants had extra time and other clinicians’ sessions were available for double coding prior to deletion. One-way, single measures, random effects intraclass correlation coefficients (ICCs) for all parent behavior and fidelity coding variables are shown in Table 1. For all fidelity variables, including the two fidelity outcomes selected for the current study, frequency of on-target comments ( $ICC = .84$ ) and percentage of on-target comments ( $ICC = .66$ ), consultants’ interrater reliability was moderate to good (Koo & Li, 2016). For parent behavior coding variables, with the exception of poor reliability on an infrequently occurring code (parental frightening, which occurred about once every 10 sessions coded,  $ICC = .13$ ), consultants’ interrater reliability was also moderate to good ( $ICCs$  range: .57-.85).

**Working Alliance**—Clinician-consultant working alliance was measured using the Working Alliance Inventory - Trainee and Supervisor Versions (WAI-T and WAI-S; Bahrck, 1989). These 36-item self-report scales are designed to measure supervisory working alliance from both the supervisor and trainee perspectives, and include items such as, “I have doubts about what we are trying to accomplish in supervision.” Respondents make ratings on 7-point Likert scales, with some items reverse-scored, and scores are averaged, with higher scores indicating more positive working alliance. The WAI-T and WAI-S have been validated through associations with supervisors’ multicultural competence and ethical behaviors, and trainees’ satisfaction with supervision (Bhat & Davis, 2007; Inman, 2006; Ladany et al., 1999). The measures include three subscales (goal agreement, task agreement, and emotional bond), but in the current study, items were averaged into a unidimensional construct, following precedent (Bhat & Davis, 2007; Inman, 2006). In the current sample, as in prior work (Bhat & Davis, 2007; Inman, 2006), internal consistency of the WAI-T and WAI-S was good (Cronbach’s alpha = .93 and .91, respectively). Both clinicians and

fidelity coding consultants rated alliance at the conclusion of fidelity-focused consultation, only. Clinicians' perceptions of Ph.D.-level clinical ABC consultants were not included in the current study, as these relationships were expected to relate less proximally to clinicians' fidelity and fidelity coding, based on previous findings that clinicians' fidelity did not change during group clinical consultation (Caron & Dozier, 2019). When clinicians received consultation from multiple fidelity coding consultants, scores were averaged to create one WAI-T score ( $M = 6.30$ ,  $SD = 0.49$ ) and one WAI-S score ( $M = 6.11$ ,  $SD = 0.54$ ) per clinician.

**Consultation Delivery**—Dosage of consultation was assessed by total number of fidelity-focused consultation sessions ( $M = 32.8$  sessions,  $SD = 8.0$ ). The number of consultation sessions prior to beginning to implement ABC was also examined ( $M = 2.4$ ,  $SD = 2.5$ ).

## Analyses

**Data Cleaning**—To create time variables for longitudinal analyses, ABC session dates were taken from fidelity coding sheets, and used to calculate months since the clinician's first fidelity-focused consultation session. Missing session dates ( $n = 47$  sessions from 8 clinicians) were estimated using the dates of coding completion and the consultation session. Some clinicians had gaps in their delivery of ABC (e.g., due to delays between completing one case and recruiting another, or session cancellations). Following procedures in Caron and Dozier (2019), these gaps in ABC delivery were represented in the data by coding periods longer than 1 month as 1-month gaps, because clinicians' fidelity was not expected to improve during long periods when they were not practicing ABC. Twenty clinicians had no gaps in implementation longer than 1 month, and some session dates were re-coded for the remaining 9 clinicians.

**Clinicians' Coding Accuracy**—Clinicians' coding accuracy was conceptualized as their interrater reliability with consultant ratings of the same session video clips; we conceptualized consultant ratings as the gold standard of "truth." We used reliability to conceptualize accuracy, as opposed to mean differences, because interrater reliability (e.g., ICCs) is the standard method for assessing accuracy of expert ratings, and ICCs offer a standardized unit of measurement across which different scales can be averaged. However, we also provide mean scores for each coding variable for clinicians and consultants, for descriptive purposes, which allows preliminary examination of mean differences. Clinicians' coding was available for 1044 (98%) of the consultant-coded sessions. Clinicians' reliability was examined across the full group using one-way, random effects ICCs. In addition to group-level ICCs, individuals' coding accuracy was also examined using one-way, single measures, random effects ICCs. Similar to correlation coefficients, ICCs cannot be calculated based on single pairs of observations because they represent the relationship between two columns of numbers (in this case, the Excel-calculated frequency of each parent behavior and fidelity statistic coded across the 5-minute segment, by both the clinician and consultant). Thus, to examine change in clinicians' coding accuracy across time, data were split into four 2.5-month time periods. The decision about time periods was based on the average duration of consultation (10.8 months), and the desire to examine change across multiple periods while retaining a sufficient session-level sample size to

estimate individual clinicians' coding accuracy within each time period, and followed a similar approach to prior work examining clinician coding reliability over time (Hogue et al., 2021). Because at very low sample sizes, one observation can greatly affect reliability statistics, individuals' coding accuracy statistics for a given time period were excluded if they were based on fewer than 5 data points, resulting in 0 missing in period 1, 4 missing in period 2, 5 missing in period 3, and 15 missing in period 4. ICCs for the six behavior coding variables and the four comment coding variables were averaged to create summary measures representing clinicians' overall behavior coding and overall comment coding accuracy.

Clinicians' growth in coding accuracy across time was examined using hierarchical linear growth models in which observations were nested within individuals. The summary measures of coding accuracy during the first time period (i.e., first 2.5 months of consultation) were used as predictors of fidelity growth. This time period was chosen because temporal precedence is important in work on predictors of change (Kazdin & Nock, 2003), and though we were unable to establish clear temporal precedence of clinicians' coding accuracy prior to fidelity, this time period provided the best available early measure of coding accuracy.

**Linear Growth Models**—Hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) was used to model the growth of coding accuracy and fidelity over time. HLM allowed a flexible approach to a dataset with a varying number of observations spaced at different intervals (level 1), nested within clinicians (level 2). Because 45% of clinicians received fidelity-focused consultation from multiple fidelity coding consultants, the model did not nest clinicians within a third level of consultants. Linear growth models were specified with the following form:  $Fidelity_{it}$  or  $CodingAccuracy_{it} = \beta_{00} + \beta_{10} * Time_{it} + r_{0i} + r_{1i} * Time_{it} + e_{it}$ . Predictors of coding accuracy or fidelity (e.g., alliance, participation in multiple baseline study) were explored by adding level 2 predictors to the growth model at both the intercept (“+  $\beta_{0i} * Predictor_i$ ”) and slope (“+  $\beta_{1i} * Predictor_i * Time_{it}$ ”). For coding accuracy growth models, time was coded -3, -2, -1, 0, such that the model intercept would represent clinicians' estimated ICCs in the final months of consultation. For fidelity growth models, time was modeled as a continuous variable (months since starting fidelity-focused consultation), and predictors were first examined in two groups, coding reliability and alliance/consultation dosage predictors, and then significant predictors that emerged were combined in simplified final models. Results presented include asymptotic (rather than robust) standard errors, which are appropriate for analyses with small sample sizes. No adjustments were made to reduce likelihood of Type I errors.

## Results

### Clinicians' Coding Accuracy

As shown in Table 1, which provides descriptive statistics about clinicians' coding accuracy, mean differences between coding scores for clinicians and consultants generally were small (i.e., effect sizes under 0.22), with the exception of clinicians under-coding parental delight ( $d = 0.48$ ) and their missed opportunities to comment ( $d = 0.56$ ). The similarity of clinicians' and consultants' mean scores across the full sample reinforced the decision to focus on

clinicians' coding accuracy using ICCs. As shown in Table 1, the group-level ICCs show significant variability in clinician and consultant coding reliability that is obscured when examining mean differences. For example, although mean differences between clinicians and consultants were similarly small for coding of parental following the lead and not following the lead behaviors, clinician and consultant interrater reliability statistics ranged from .44 - .89 for these codes. Of note, however, by the final period of fidelity-focused consultation (i.e., Training Period 4), clinicians' coding of many variables was on par with consultants' interrater reliability statistics.

Next, we tested whether clinicians demonstrated growth in coding accuracy across time. Table 2 shows the results of hierarchical linear growth models estimating growth in clinicians' coding accuracy (i.e., individual-level ICCs) of different variables over time. As hypothesized, clinicians' coding accuracy for four of the six behavior coding variables, as well as their average behavior coding accuracy and average comment coding accuracy, increased over time. As reflected by the coefficients from the models in Table 2 that reached significance, clinicians' ICCs were expected to increase by .04 to .09 (on a scale from 0.00 to 1.00) every 2.5 months.

We explored whether study group (i.e., participation in the multiple baseline study for 6 clinicians) affected results by adding a study group predictor at the intercept and slopes for each of the coding accuracy growth models. The coding accuracy of multiple baseline clinicians differed from the rest of the sample only in demonstrating poorer initial accuracy for coding not following the lead behaviors ( $\beta_{01} = -0.30, p = .032$ ), and inclusion of the study group variable did not change the results in Table 2 (i.e., all 6 coding accuracy variables that showed significant growth continued to show significant growth even when accounting for effect of study group at the intercept and slope). Because differences in the models were minimal, we present the simpler models in Table 2.

### Clinicians' Fidelity

Clinicians' fidelity statistics, as coded by consultants and averaged across the full training year, are shown in the bottom half of the third column of Table 1. On average, clinicians met ABC certification criteria: frequency of >1.0 comments/minute or <50% missed opportunities to comment, >80% on-target comments, and >1.0 component per comment. Clinicians' growth in fidelity was examined next. As shown by the significant positive slopes in the first two models in Table 3, clinicians' comment frequency and percentage of on-target comments increased over the consultation period. The first two models in Table 3 also demonstrate that participation in the multiple baseline study was associated with poorer initial fidelity but greater growth over time, so this variable was included in subsequent models.

**Coding Accuracy Predictors of Fidelity**—Next, the two summary coding measures, behavior coding accuracy and comment coding accuracy, were examined as predictors of initial fidelity and growth over time. As shown in Table 3 (models 3-4), behavior coding accuracy was associated with both model intercepts, such that clinicians who were initially better at coding parent behaviors were able to comment more frequently and have

their comments be more often on-target when they began implementing ABC. Comment coding accuracy was negatively associated with the comment frequency intercept, such that clinicians who were initially better at coding their own fidelity made less frequent comments when they began implementing ABC. However, comment coding accuracy was positively associated with the slope of percentage of on-target comments, such that clinicians who were initially better at coding their own comments showed stronger growth in their percentage of on-target comments over time than clinicians who were weaker at coding initially.

**Other Predictors of Fidelity**—Next, clinician-consultant working alliance and consultation dosage variables were tested as predictors of the intercept and slope of the fidelity outcomes. Table 3 presents these results in models 5-6. Higher consultant-reported working alliance was associated with clinicians making more frequent comments when they began implementing ABC. In addition, the number of consultation sessions that clinicians attended prior to starting to implement ABC was associated with stronger growth in comment frequency over time, suggesting that there was a benefit to meeting with a fidelity coding consultant even before starting to code one's own sessions. None of the alliance or consultation dosage variables were associated with initial levels or growth over time of percentage of on-target comments.

Finally, the significant coding accuracy, alliance, and consultation dosage predictors of comment frequency identified in the previous models were combined (model 7). As shown at the bottom of Table 3, clinicians' early behavior coding accuracy and consultant-reported working alliance continued to be positively associated with initial frequency of making in-the-moment comments. The number of sessions of consultation prior to beginning ABC implementation and clinicians' comment coding accuracy were no longer significantly associated with the slope or intercept.

### **Predictors of Growth in Self-Coding Accuracy**

Finally, predictors of growth in self-coding accuracy were examined. As shown in Table 4, one significant predictor emerged. Higher clinician-reported working alliance was associated with lower estimated behavior coding accuracy (i.e., model intercept) at the end of training. It appeared that this may have been driven by borderline slower growth in clinicians' behavior coding accuracy ( $\beta_{11} = -0.07, p = .053$ ). In fact, when time was modeled so the intercept was in the first coding period, clinician-reported working alliance was no longer associated with a lower intercept for behavior coding accuracy (Coefficient = 0.00,  $t = 0.05, p = .96$ ), further suggesting that the impact of clinician-reported working alliance accumulated across training due to an association with slower growth. As with the earlier coding accuracy models shown in Table 2, we tested whether adding participation in the multiple baseline design would affect results, and because it did not (and was not a significant predictor of coding accuracy outcomes), we present simpler models here.

## **Discussion**

The current study examined clinicians' growth in ABC fidelity and self-coding accuracy during consultation that involved completing self-coding and receiving consultant coding-

based fidelity feedback. The results build on the work of Caron and Dozier (2019) by replicating in a larger sample the findings that clinicians' frequency of making in-the-moment comments and percentage of on-target comments increased during a year of fidelity-focused consultation. In addition, as hypothesized, the current study demonstrated that clinicians' ability to code their own ABC fidelity improved during consultation, a novel finding. Together, these findings suggest that fidelity-focused consultation may support clinicians' understanding of intervention fidelity, as well as their actual growth in fidelity over time.

This study is one of the first to demonstrate improvement in clinicians' self-coding reliability during training or consultation, and the first to our knowledge outside the field of cognitive behavioral therapy. Because clinicians' self-rating ability is often viewed as poor (e.g., Martino et al., 2009), it is important to emphasize that by the end of consultation, clinicians demonstrated moderate to excellent reliability on 8 of the 10 behavior and comment coding variables at the group level. Certain aspects of the ABC fidelity measure may enhance clinicians' ability to self-code accurately; specifically, its focus on explicit verbal behaviors may be less subjective and prone to self-enhancement biases than global rating scales. However, because self-coding occurred only in the context of consultation, the current results may not generalize to independent self-coding. Specifically, because consultants reviewed clinicians' coding and provided feedback on inaccurate coding, clinicians likely were incentivized to invest time and attention and to complete coding accurately. In addition, because clinicians in this study made ratings using self-observation of video, the current results should not be expected to generalize to clinicians' retrospective self-report, a more common fidelity assessment technique. Further, more research is needed on the accessibility and sustainability of self-coding practice, particularly given the administrative burden and burnout experienced by many community-based clinicians. Although the time investment in the current study was considerable ( $\frac{1}{2}$  – 1 hour weekly to code prior to consultation, plus  $\frac{1}{2}$  hour consultation meeting time), it is possible that similar gains could be made from less frequent self-coding and/or consultation. Of note, Caron & Dozier (2019) found that clinicians who completed a year of fidelity-focused consultation maintained fidelity for up to two years following the end of consultation. Beale et al. (2020) similarly found that a subset of clinicians who participated in follow up maintained both self-coding accuracy and fidelity one year after the conclusion of training. These findings suggest that learning to self-code may allow clinicians to accurately self-monitor in vivo and to maintain fidelity during a sustainment period, potentially even without actually self-coding.

In addition to finding that both clinicians' fidelity and their fidelity self-coding improved during consultation, the current study also identified associations between clinicians' coding accuracy and their fidelity. First, clinicians' accuracy in coding parent behaviors was positively associated with their initial frequency of in-the-moment commenting and percentage of on-target comments. To make in-the-moment comments, clinicians must quickly and correctly identify parent behaviors in session. Clinicians who are able to code parent behaviors correctly when reviewing session video are likely more skilled at identifying parent behaviors in session, allowing them to more easily make comments, and to make comments that are accurate and appropriate. Though this study did not find



that clinicians' coding of client behavior was linked to *growth* in fidelity, early focus in consultation on clinicians' accuracy of client behavior coding should be examined as a potential active ingredient of consultation that may lead to improvements in fidelity. Coding of client behaviors is also used in a limited number of other fidelity assessments, such as the coach coding system for Parent-Child Interaction Therapy (Barnett et al., 2014) and the Motivational Interviewing Consultation and Feedback Form (Isenhart et al., 2014). Isenhart et al. (2014) theorized that coding client behaviors such as change talk would increase clinicians' awareness of these behaviors in their own sessions, and allow clinicians to respond appropriately when they occurred. The current study provides evidence for the link between clinicians' awareness of relevant client behavior and their ability to respond appropriately.

In addition, the current study found links between clinicians' early accuracy of coding their own comments and their initial fidelity and growth in fidelity over time. Surprisingly, clinicians' early accuracy in coding their own fidelity was negatively associated with initial comment frequency. It may be that greater early understanding of fidelity coding may initially interfere with making frequent comments because it leads clinicians to over-think the process of making comments as they evaluate how their hypothetical comment would be scored, and to miss the opportunity to say something before the parent's behavior changes. Somewhat relatedly, McManus et al. (2012) found that more competent trainees had greater self-rating discrepancies from supervisors than less competent trainees, and Hitzeman et al. (2020) found that trainees' reflective practice ability was positively associated with the tendency to underestimate competence. Together, these findings suggest that associations between self-rating ability and fidelity may be complex, particularly among new trainees for whom high levels of motivation and introspection may be accompanied by anxiety and self-doubt (Hitzeman et al., 2020).

However, less surprisingly, clinicians' early accuracy in coding their own fidelity was positively associated with growth in percentage of on-target comments. That is, clinicians' understanding of the component aspects of ABC fidelity (e.g., what statements count as in-the-moment comments, whether comments are on- or off-target) predicted growth in their fidelity over time. These results support Isenhart et al. (2014)'s assertion that engaging in fidelity coding trains clinicians to recognize intervention-consistent responses to clients, and translates to actually making intervention-consistent responses to clients in session. Further, the results build on prior findings that more competent clinicians tend to be more accurate self-raters than less competent clinicians (e.g., Beale et al., 2020) and suggest that understanding of fidelity, as reflected by clinicians' ability to code fidelity accurately, may actually be a *precursor* to increasing some aspects of fidelity.

With regard to working alliance and consultation dosage as predictors of fidelity, there was limited evidence that these constructs were active ingredients of consultation in the current study. However, findings illuminate factors that are important to consider and measure in implementation contexts. For example, working alliance, as reported by consultants, was linked to higher clinician fidelity at the beginning of consultation, suggesting that clinicians who make more frequent in-the-moment comments may be perceived by consultants as easier to train and work with than clinicians who struggle more with ABC fidelity. Although

clinician-consultant working alliance has not consistently been linked to implementation outcomes (Schoenwald et al., 2004; Wehby et al., 2012), this association may be important to consider if use of fidelity coding-based feedback becomes more common in consultation protocols.

One result related to working alliance initially seemed surprising. Specifically, higher clinician-reported working alliance was associated with lower coding accuracy at the end of training, likely driven by slower growth in coding accuracy. Perhaps it is possible for clinicians to feel too comfortable in the consultation working relationship, and not feel compelled to change. Though surprising, this result is not unprecedented in the implementation literature, as Schoenwald et al. (2004) found that stronger clinician-reported alliance with consultants was associated with poorer fidelity and client outcomes. However, these results are tempered by the limitation that working alliance was measured only at the end of consultation, preventing temporal interpretation of associations, an important aspect of active ingredients research. In addition, the fidelity coding consultants' status as primarily undergraduate students may have influenced these results in a way that would likely not generalize to many implementation contexts.

On the whole, there is one clear recommendation for implementation practices from the consultation dosage predictors examined in the current study. When clinicians do not begin implementing an evidence-based intervention immediately after training, they are likely to benefit from consultation in the period before beginning to use the intervention with clients. Specifically, in the current study, higher frequency of fidelity-focused consultation sessions prior to the first ABC session, which ranged from 0 to 8, was associated with stronger growth in comment frequency over time, suggesting that when clinicians did not immediately begin ABC and received consultation during the delay, they were able to show more rapid growth once they did begin implementing ABC. In fidelity-focused consultation, when there is no session video or coding to review, consultants use active learning activities, such as worksheets and videos that give clinicians opportunities to practice recognizing and coding both parent behaviors and in-the-moment comments, and to practice making in-the-moment comments. These activities may lay the groundwork for future growth in fidelity.

### Strengths and Limitations

The impact of the current results is bounded by several limitations. First, the sample size of 29 was fairly small, though larger than some other studies that have examined self-coding of fidelity in the context of consultation ( $N = 7$  in Caron & Dozier, 2019;  $N = 2$  in Isenhardt et al., 2014). The relatively large number of consultants ( $N = 22$ ) is a strength, but at the same time, leads to near singularity with clinicians, making it impossible to disentangle sources of variance attributable to each. Further, the fact that 45% of the sample received consultation from two or three fidelity coding consultants prevented modeling of clinicians nested within consultants, and also resulted in working alliance scores that were averaged across multiple fidelity coding consultants and did not capture differences in working relationships that varied in length and timing during the year of consultation. It is also possible that the four clinicians who declined to participate felt less positive about

or comfortable with fidelity-focused consultation, and their exclusion may have skewed results. Given the preliminary nature of the current study, we did not adjust the p-values in our multiple analyses to account for increased likelihood of type 1 errors, and the possibility of type 1 errors should be noted. In addition, training and supervision of the fidelity coding consultants was considerable, increasing internal validity, but presenting a possible limitation to external validity and replicability. Despite this training, consultants' interrater reliability on fidelity coding variables ranged from .66 to .84, and their interrater reliability on parent behavior coding variables ranged from .13 to .85. Further, interrater reliability between consultants was evaluated from only 6% of the sample, as collection of sessions for purposes of coding interrater reliability was limited by implementation constraints (i.e., limited consultant time and deleting sessions quickly to protect client confidentiality). Therefore, consultants' coding has limitations in its treatment as the gold standard in this study. Nevertheless, consultant or observer coding of clinician fidelity with similar or lower interrater reliability has been used in other studies (e.g., Brookman-Frazer et al., 2021; Masia Warner et al., 2013; McManus et al., 2012; Stice et al., 2013).

Most importantly, data are correlational. Although growth in coding ability was observed over the course of the study, experimental designs are needed to establish fidelity-focused consultation as a *cause* of these improvements, since clinicians were also gaining additional experience implementing ABC and receiving group clinical ABC consultation over the course of the study. Further, a trial in which clinicians were randomized to fidelity-focused consultation with and without self-coding would be the only way to definitively demonstrate that the process of self-coding (versus just receiving consultants' fidelity feedback) causes improvement in fidelity, and therefore validate its role as an active ingredient of consultation. Future work should include measures of potential mediators of effects, such as effort, reflective practice ability, self-efficacy, and receptivity to consultants' fidelity feedback, which were not included in the current study, and should measure these constructs at various timepoints in training/consultation, rather than at the end of the consultation period. In addition, follow-up work could explore temporal patterns of change in both coding ability and fidelity (e.g., do they change simultaneously, or does improvement in coding ability precede growth in fidelity?). Further, other models of support for fidelity coding and fidelity should be explored, including group-based fidelity-focused consultation, less frequent dosages of consultation, and models of self-guided learning (e.g., learning to code fidelity from video examples with gold standard codes; Hogue et al., 2021), as these models could provide similar benefits with lower implementation costs. The current results may not generalize outside of ABC implementation, although there are promising signals in the field of cognitive behavioral therapy (Loades & Myles, 2016; Beale et al., 2020), so examination of self-coding in the context of consultation for various other interventions is a critical next step.

The study's limitations are countered by several strengths. First, although the clinician sample size was relatively small, the session-level sample size was much larger, and allowed for nuanced modeling of change in fidelity and fidelity coding accuracy across time. The fact that data were gathered in an implementation context, as opposed to supervision during a randomized controlled trial, increases potential generalizability to other implementation

contexts. In addition, clinicians worked in a variety of agencies in six US states, increasing generalizability to some extent.

The greatest strength of the current study lies in its novel findings linking self-coding accuracy with growth in fidelity. These findings offer several important, generalizable recommendations. First, in designing fidelity measures for consultation, intervention developers and implementation scientists should consider incorporating measurement of client behaviors that serve as triggers for key clinician responses. In the current study, accurate recognition of relevant client behaviors was linked to initial clinician fidelity, suggesting that including client behaviors in fidelity assessments may improve clinicians' ability to recognize these behaviors in session and respond appropriately. Second, fidelity feedback is becoming increasingly common in evidence-based consultation protocols (e.g., Eiraldi et al., 2018; Liness et al., 2019; Martino et al., 2016; Weck et al., 2017). The current study suggests that incorporating clinician self-coding of fidelity into consultation may result in additional, unique benefits. Consultants should provide clinicians with coding training and feedback about their self-coding accuracy, as clinicians' ability to accurately code their own fidelity may be linked to their later growth in fidelity. In sum, this study provides preliminary support for self-coding of fidelity as a potential active ingredient of consultation with links to improved fidelity. Understanding active ingredients of consultation will facilitate development of implementation protocols that are more effective in changing clinician behavior and promoting client outcomes. However, these findings are novel, the study is correlational, and generalizability is unknown, so additional research is needed.

## Funding

This study was supported by the National Institutes of Health under Grants R01 MH052135, R01 MH074374, and R01 MH084135.

We would like to thank the parent coaches and consultants who participated in this study.

## References

- Bahrack AS (1989). Role induction for counselor trainees: Effects on the supervisory working alliance (Doctoral dissertation). Retrieved from OhioLINK Electronic Theses & Dissertations Center.
- Barnett ML, Niec LN, & Acevedo-Polakovich ID (2014). Assessing the key to effective coaching in parent-child interaction therapy: The therapist-parent interaction coding system. *Journal of Psychopathology and Behavioral Assessment*, 36(2), 211–223. 10.1007/s10862-013-9396-8 [PubMed: 24839350]
- Beale S, Liness S, & Hirsch CR (2020). Trainee self-assessment of cognitive behaviour therapy competence during and after training. *The Cognitive Behaviour Therapist*, 13. 10.1017/S1754470X19000357
- Bearman SK, Schneiderman RL, & Zoloth E (2017). Building an evidence base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(2), 293–307. 10.1007/s10488-016-0723-8 [PubMed: 26867545]
- Bearman SK, Weisz JR, Chorpita BF, Hoagwood K, Ward A, Ugueto AM, Bernstein A & The Research Network on Youth Mental Health. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(6), 518–529. 10.1007/s10488-013-0485-5 [PubMed: 23525895]

- Beidas RS, Edmunds JM, Marcus SC, & Kendall PC (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services*, 63(7), 660–665. 10.1176/appi.ps.201100401 [PubMed: 22549401]
- Bennett-Levy J, Thwaites R, Chaddock A, & Davis M (2009). Reflective practice in cognitive behavioural therapy: The engine of lifelong learning. In Stedmon J & Dallos R (Eds.), *Reflective practice in psychotherapy and counselling*, 115–135. McGraw Hill.
- Bernard K, Dozier M, Bick J, & Gordon MK (2015). Intervening to enhance cortisol regulation among children at risk for neglect: results of a randomized clinical trial. *Development and Psychopathology*, 27(3), 829–841. 10.1017/S095457941400073X [PubMed: 25156798]
- Bernard K, Dozier M, Bick J, Lewis-Morrarty E, Lindhiem O, & Carlson E (2012). Enhancing attachment organization among maltreated children: Results of a randomized clinical trial. *Child Development*, 83(2), 623–636. 10.1111/j.1467-8624.2011.01712.x [PubMed: 22239483]
- Bhat CS, & Davis TE (2007). Counseling supervisors' assessment of race, racial identity, and working alliance in supervisory dyads. *Journal of Multicultural Counseling and Development*, 35(2), 80–91. 10.1002/j.2161-1912.2007.tb00051.x
- Brookman-Frazer L, Stadnick NA, Lind T, Roesch S, Terrones L, Barnett ML, Regan J, Kennedy CA, Garland AF, & Lau AS (2021). Therapist-observer concordance in ratings of EBP strategy delivery: Challenges and targeted directions in pursuing pragmatic measurement in children's mental health services. *Administration and Policy in Mental Health and Mental Health Services Research*, 48, 155–170. 10.1007/s10488-020-01054-x [PubMed: 32507982]
- Brosan L, Reynolds S, & Moore RG (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy*, 36(5), 581–587. 10.1017/s1352465808004438
- Caron E. (2017). Effects and processes of fidelity-focused consultation (Doctoral dissertation, University of Delaware). UDSpace Digital Archive. [https://udspace.udel.edu/bitstream/handle/19716/23096/Caron\\_udel\\_0060D\\_13001.pdf?sequence=1](https://udspace.udel.edu/bitstream/handle/19716/23096/Caron_udel_0060D_13001.pdf?sequence=1)
- Caron E, Bernard K, & Dozier M (2018). In vivo feedback predicts parent behavior change in the Attachment and Biobehavioral Catch-up intervention. *Journal of Clinical Child & Adolescent Psychology*, 47(sup1), S35–S46. 10.1080/15374416.2016.1141359 [PubMed: 27043449]
- Caron E, & Dozier M (2019). Effects of fidelity-focused consultation on clinicians' implementation: An exploratory multiple baseline design. *Administration and Policy in Mental Health and Mental Health Services Research*, 46(4), 445–457. 10.1007/s10488-019-00924-3 [PubMed: 30783903]
- Caron E, Lind TA, & Dozier M (2021). Strategies that promote therapist engagement in active and experiential learning: micro-level sequential analysis. *The Clinical Supervisor*, 40(1), 112–133. 10.1080/07325223.2020.1870023 [PubMed: 34248258]
- Caron E, Muggeo MA, Souer HR, Pella JE, & Ginsburg GS (2020). Concordance between clinician, supervisor and observer ratings of therapeutic competence in CBT and treatment as usual: does clinician competence or supervisor session observation improve agreement?. *Behavioural and Cognitive Psychotherapy*, 48(3), 350–363. 10.1017/s1352465819000699 [PubMed: 31806076]
- Carroll K, Nich C, & Rounsaville B (1998). Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research*, 8(3), 307–320.
- Child and Adolescent Health Measurement Initiative. 2018-2019 National Survey of Children's Health (NSCH) data query. Data Resource Center for Child and Adolescent Health supported by the U.S. Department of Health and Human Services, Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB). Retrieved 05/03/21 from [www.childhealthdata.org](http://www.childhealthdata.org).
- Collyer H, Eisler I, & Woolgar M (2020). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child & Adolescent Psychiatry*, 29(4), 417–431. [PubMed: 30604132]
- Cox JR, McLeod BD, Jensen-Doss A, Srivastava V, Southam-Gerow MA, Kendall PC, & Weisz JR (2020). Examining how CBT interventions for anxious youth are delivered across settings. *Behavior Therapy*, 51(6), 856–868. 10.1016/j.beth.2019.11.008 [PubMed: 33051029]
- Dozier M, Bernard K, & Roben CKP (2017). Attachment and biobehavioral catch-up. In Steele H & Steele M (Eds.), *The handbook of attachment-based interventions*, 27–49. The Guilford Press.

- Dunn C, Darnell D, Atkins DC, Hallgren KA, Imel ZE, Bumgardner K, Owens M, & Roy-Byrne P (2016). Within-provider variability in motivational interviewing integrity for three years after MI training: does time heal?. *Journal of Substance Abuse Treatment*, 65, 74–82. 10.1016/j.jsat.2016.02.008 [PubMed: 27016875]
- Edmunds JM, Beidas RS, & Kendall PC (2013). Dissemination and implementation of evidence-based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice*, 20(2), 152–165. 10.1111/cpsp.12031 [PubMed: 24072959]
- Edmunds JM, Kendall PC, Ringle VA, Read KL, Brodman DM, Pimentel SS, & Beidas RS (2013). An examination of behavioral rehearsal during consultation as a predictor of training outcomes. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(6), 456–466. 10.1007/s10488-013-0490-8 [PubMed: 23616234]
- Eiraldi R, Mautone JA, Khanna MS, Power TJ, Orapallo A, Cacia J, Schwartz BS, McCurdy B, Keiffer J, Paidipati C, Kanine R, Abraham M, Tulio S, Swift L, Bressler SN, Cabello B, & Jawad AF (2018). Group CBT for externalizing disorders in urban schools: Effect of training strategy on treatment fidelity and child outcomes. *Behavior Therapy*, 49(4), 538–550. 10.1016/j.beth.2018.01.001 [PubMed: 29937256]
- Fixsen DL, Naoom SF, Blase KA, Friedman RM, & Wallace F (2005). Implementation research: A synthesis of the literature (FMHI Publication No. 231). University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Funderburk B, Chaffin M, Bard E, Shanley J, Bard D, & Berliner L (2015). Comparing client outcomes for two evidence-based treatment consultation strategies. *Journal of Clinical Child & Adolescent Psychology*, 44(5), 730–741. 10.1080/15374416.2014.910790 [PubMed: 24871692]
- Garnett M, Bernard K, Hoye J, Zajac L, & Dozier M (2020). Parental sensitivity mediates the sustained effect of Attachment and Biobehavioral Catch-up on cortisol in middle childhood: A randomized clinical trial. *Psychoneuroendocrinology*, 121, 104809. 10.1016/j.psyneuen.2020.104809 [PubMed: 32781397]
- Hitzeman C, Gonsalvez CJ, Britt E, & Moses K (2020). Clinical psychology trainees' self versus supervisor assessments of practitioner competencies. *Clinical Psychologist*, 24(1), 18–29. 10.1111/cp.12183
- Hogue A, Dauber S, Lichvar E, Bobek M, & Henderson CE (2015). Validity of therapist self-report ratings of fidelity to evidence-based practices for adolescent behavior problems: Correspondence between therapists and observers. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(2), 229–243. 10.1007/s10488-014-0548-2 [PubMed: 24711046]
- Hogue A, Porter N, Bobek M, MacLean A, Bruynesteyn L, Jensen-Doss A, ... & Henderson CE (2021). Online training of community therapists in observational coding of family therapy techniques: Reliability and accuracy. *Administration and Policy in Mental Health and Mental Health Services Research*. Advance online publication. 10.1007/s10488-021-01152-4
- Hulleman CS, & Cordray DS (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88–110. 10.1080/19345740802539325
- Inman AG (2006). Supervisor multicultural competence and its relation to supervisory process and outcome. *Journal of Marital and Family Therapy*, 32(1), 73–85. 10.1111/j.1752-0606.2006.tb01589.x [PubMed: 16468682]
- Institute of Medicine (2015). Psychosocial interventions for mental and substance use disorders. National Academies Press. 10.17226/19013.
- Isenhardt C, Dieperink E, Thuras P, Fuller B, Stull L, Koets N, & Lenox R (2014). Training and maintaining motivational interviewing skills in a clinical trial. *Journal of Substance Use*, 19(1-2), 164–170. 10.3109/14659891.2013.765514
- Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, & Hofmann SG (2016). Quantity and quality of homework compliance: A meta-analysis of relations with outcome in cognitive behavior therapy. *Behavior Therapy*, 47(5), 755–772. [PubMed: 27816086]
- Kazdin AE, & Nock MK (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44(8), 1116–1129. 10.1111/1469-7610.00195 [PubMed: 14626454]



- Koo TK, & Li MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. 10.1016/j.jcm.2016.02.012 [PubMed: 27330520]
- Ladany N, Lehrman-Waterman D, Molinaro M, & Wolgast B (1999). Psychotherapy supervisor ethical practices: Adherence to guidelines, the supervisory working alliance, and supervisee satisfaction. *The Counseling Psychologist*, 27(3), 443–475. 10.1177/0011000099273008
- Lind T, Bernard K, Yarger HA, & Dozier M (2020). Promoting compliance in children referred to child protective services: A randomized clinical trial. *Child Development*, 91(2), 563–576. 10.1111/cdev.13207 [PubMed: 30815861]
- Lind T, Raby KL, Caron E, Roben CK, & Dozier M (2017). Enhancing executive functioning among toddlers in foster care with an attachment-based intervention. *Development and Psychopathology*, 29(2), 575–586. 10.1017/S0954579417000190 [PubMed: 28401847]
- Liness S, Beale S, Lea S, Byrne S, Hirsch CR, & Clark DM (2019). The sustained effects of CBT training on therapist competence and patient outcomes. *Cognitive Therapy and Research*, 43(3), 631–641.
- Loades ME, & Myles PJ (2016). Does a therapist's reflective ability predict the accuracy of their self-evaluation of competence in cognitive behavioural therapy? *The Cognitive Behaviour Therapist*, 9. 10.1017/S1754470X16000027
- Magill M, Kiluk BD, McCrady BS, Tonigan JS, & Longabaugh R (2015). Active ingredients of treatment and client mechanisms of change in behavioral treatments for alcohol use disorders: Progress 10 years later. *Alcoholism: Clinical and Experimental Research*, 39(10), 1852–1862. 10.1111/acer.12848
- Martino S, Ball S, Nich C, Frankforter TL, & Carroll KM (2009). Correspondence of motivational enhancement treatment integrity ratings among therapists, supervisors, and observers. *Psychotherapy Research*, 19(2), 181–193. 10.1080/10503300802688460 [PubMed: 19396649]
- Martino S, Paris M, Añez L, Nich C, Canning-Ball M, Hunkele K, Olmstead TA, & Carroll KM (2016). The effectiveness and cost of clinical supervision for motivational interviewing: a randomized controlled trial. *Journal of Substance Abuse Treatment*, 68, 11–23. 10.1016/j.jsat.2016.04.005 [PubMed: 27431042]
- Masia Warner C, Brice C, Esseling PG, Stewart CE, Mufson L, & Herzig K (2013). Consultants' perceptions of school counselors' ability to implement an empirically-based intervention for adolescent social anxiety disorder. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(6), 541–554. 10.1007/s10488-013-0498-0 [PubMed: 23716144]
- Mathieson FM, Barnfield T, & Beaumont G (2009). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist*, 2(1), 43–50. 10.1017/S1754470X08000081
- McLeod BD, Cox JR, Jensen-Doss A, Herschell A, Ehrenreich-May J, & Wood JJ (2018). Proposing a mechanistic model of clinician training and consultation. *Clinical Psychology: Science and Practice*, 25(3), e12260. 10.1111/cpsp.12260 [PubMed: 30713369]
- McManus F, Rakovshik S, Kennerley H, Fennell M, & Westbrook D (2012). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology*, 51(3), 292–306. 10.1111/j.2044-8260.2011.02028.x
- Moyers TB, Manuel JK, Wilson PG, Hendrickson SM, Talcott W, & Durand P (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, 36(2), 149–162. 10.1017/S1352465807004055
- Niehaus E, Campbell CM, & Inkelas KK (2014). HLM behind the curtain: Unveiling decisions behind the use and interpretation of HLM in higher education research. *Research in Higher Education*, 55(1), 101–122. 10.1007/s11162-013-9306-7
- Peavy KM, Guydish J, Manuel JK, Campbell BK, Lisha N, Le T, Delucchi K & Garrett S (2014). Treatment adherence and competency ratings among therapists, supervisors, study-related raters and external raters in a clinical trial of a 12-step facilitation for stimulant users. *Journal of Substance Abuse Treatment*, 47(3), 222–228. 10.1016/j.jsat.2014.05.008 [PubMed: 25017448]
- Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, & Mittman B (2009). Implementation research in mental health services: an emerging science with conceptual, methodological, and

- training challenges. *Administration and Policy in Mental Health and Mental Health Services Research*, 36, 24–34. 10.1007/s10488-008-0197-4 [PubMed: 19104929]
- Raby KL, Freedman E, Yarger HA, Lind T, & Dozier M (2019). Enhancing the language development of toddlers in foster care by promoting foster parents' sensitivity: Results from a randomized controlled trial. *Developmental Science*, 22(2), e12753. 10.1111/desc.12753 [PubMed: 30230658]
- Raudenbush SW, & Bryk AS (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Reichelt FK, James IA, & Blackburn IM (2003). Impact of training on rating competence in cognitive therapy. *Journal of Behavior Therapy and Experimental Psychiatry*, 34(2), 87–99. 10.1016/s0005-7916(03)00022-3 [PubMed: 12899893]
- Schoenwald SK, Sheidow AJ, & Letourneau EJ (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology*, 33(1), 94–104. 10.1207/S15374424JCCP3301\_10 [PubMed: 15028545]
- Schalbe CS, Oh HY, & Zweben A (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, 109(8), 1287–1294. 10.1111/add.12558 [PubMed: 24661345]
- Serfaty M, Shafran R, Vickerstaff V, & Aspden T (2020). A pragmatic approach to measuring adherence in treatment delivery in psychotherapy. *Cognitive Behaviour Therapy*, 49(5), 347–360. 10.1080/16506073.2020.1717594 [PubMed: 32114905]
- Stice E, Butryn ML, Rohde P, Shaw H, & Marti CN (2013). An effectiveness trial of a new enhanced dissonance eating disorder prevention program among female college students. *Behaviour Research and Therapy*, 51(12), 862–871. 10.1016/j.brat.2013.10.003 [PubMed: 24189570]
- Tabak RG, Khoong EC, Chambers DA, & Brownson RC (2012). Bridging research and practice: models for dissemination and implementation research. *American Journal of Preventive Medicine*, 43(3), 337–350. 10.1016/j.amepre.2012.05.024 [PubMed: 22898128]
- Wain RM, Kutner BA, Smith JL, Carpenter KM, Hu MC, Amrhein PC, & Nunes EV (2015). Self-report after randomly assigned supervision does not predict ability to practice Motivational Interviewing. *Journal of Substance Abuse Treatment*, 57, 96–101. 10.1016/j.jsat.2015.04.006 [PubMed: 25963775]
- Weck F, Kaufmann YM, & Höfling V (2017). Competence feedback improves CBT competence in trainee therapists: A randomized controlled pilot study. *Psychotherapy Research*, 27(4), 501–509. 10.1080/10503307.2015.1132857 [PubMed: 26837800]
- Wehby JH, Maggin DM, Moore Partin TC, & Robertson R (2012). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the good behavior game. *School Mental Health*, 4, 22–33. 10.1007/s12310-011-9067-4
- Weisz JR, Kuppens S, Eckshtain D, Ugueto AM, Hawley KM, & Jensen-Doss A (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: a multilevel meta-analysis. *JAMA Psychiatry*, 70(7), 750–761. 10.1001/jamapsychiatry.2013.1176 [PubMed: 23754332]
- Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, ... & Fordwood SR (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist*, 72(2), 79–117. 10.1037/a0040360

**Table 1**  
 Clinicians' Coding Accuracy: Group Level Means and ICCs with Consultant Comparison

Variable	Ratings in 5-Minute Session Clips			Clinician ICCs during Each 2.5-Month Training Period				
	Clinician M (SD)	Consultant M (SD)	Cohen's <i>d</i>	1	2	3	4	Consultant ICC
<i>Parent Behavior Coding</i>								
Following the Lead	8.04 (6.05)	8.62 (6.16)	0.18***	.80	.87	.89	.87	<b>.80</b>
Not Following the Lead	3.52 (3.49)	3.81 (3.12)	0.09**	.44	.61	.67	.50	<b>.63</b>
Delight	3.00 (3.00)	4.21 (3.87)	0.48***	.63	.65	.77	.69	<b>.85</b>
Nurturing	0.78 (1.51)	0.79 (1.40)	0.01	.77	.76	.82	.79	<b>.57</b>
Not Nurturing	0.30 (0.85)	0.41 (1.08)	0.12***	.43	.67	.64	.80	<b>.77</b>
Frightening	0.11 (0.51)	0.09 (0.46)	-0.04	.63	.57	.18	.36	<b>.13</b>
Average Behavior Coding	--	--	--	.45	.58	.57	.63	--
<i>Comment Coding</i>								
On-Target Comment Frequency	1.36 (1.05)	1.45 (1.10)	0.22***	.91	.90	.94	.93	<b>.84</b>
Percent of On-Target Comments	86.21 (25.05)	88.89 (22.54)	0.12***	.43	.54	.58	.66	<b>.66</b>
Percent of Missed Opportunities	36.16 (27.99)	51.30 (22.39)	0.56***	.25	.30	.32	.35	<b>.72</b>
Average Components	1.32 (0.55)	1.31 (0.49)	-0.02	.53	.70	.76	.75	<b>.72</b>
Average Comment Coding	--	--	--	.40	.50	.50	.54	--

*Note.* Bold indicates the gold standard consultant coder scores; Italics are used to highlight the values of the summary measures Average Behavior Coding and Average Comment Coding Parent behavior codes represent the frequency of each parent behavior coded, and comment codes represent a summary of individually-coded comments, as calculated by an Excel coding sheet completed by clinicians and consultants while observing a 5-min session clip. Cohen's *d* represents the effect size of the difference in mean levels of clinician and consultant coding for the same session. Clinician ICCs represent clinicians' agreement with consultant coding; although nearly all ABC sessions used in consultation were coded by both the consultant and the clinician, the Single Measures ICC was used for more appropriate comparison with the Consultant ICC (for which a smaller portion of the sample was coded). Koo and Li's (2016) guidelines for interpreting interrater reliability are: <.5 = poor; .5-.74 = moderate; .75-.89 = good; >.9 = excellent

\*  $p < .05$   
 \*\*  $p < .01$   
 \*\*\*  $p < .01$

**Table 2**

Clinicians' Coding Accuracy: Linear Growth Models of Change over Time

Variable	Coefficient	SE	<i>t</i> -ratio	<i>p</i> -value	Intercept
<i>Parent Behavior Coding</i>					
Following the Lead	.05	.02	2.70	.012	.82
Not Following the Lead	.05	.03	2.11	.044	.50
Delight	.07	.03	2.52	.018	.67
Nurturing	.09	.03	3.15	.004	.84
Not Nurturing	.01	.04	0.18	.86	.46
Frightening	-.01	.05	-0.16	.88	.29
<i>Average Behavior Coding</i>	<i>.05</i>	<i>.01</i>	<i>3.64</i>	<i>.001</i>	<i>.64</i>
<i>Comment Coding</i>					
On-Target Comment Frequency	.03	.02	1.85	.08	.89
Percent of On-Target Comments	.06	.04	1.54	.14	.36
Percent of Missed Opportunities	.06	.04	1.46	.16	.35
Average Components	.02	.02	1.00	.33	.59
<i>Average Comment Coding</i>	<i>.05</i>	<i>.02</i>	<i>2.49</i>	<i>.019</i>	<i>.55</i>

*Note.* Italics are used to highlight the values of the summary measures Average Behavior Coding and Average Comment Coding Coefficients represent the expected increase in ICC score (on a scale from 0.00 to 1.00) every 2.5 months. Intercepts represent the model-estimated ICCs in the final 2.5 months of training. These growth models examine change in clinicians' coding accuracy over time at the individual level, whereas the ICCs in Table 1 are calculated across the full group.

**Table 3**

Predictors of Initial Fidelity and Fidelity Growth over the Consultation Period

Effect	Coefficient	SE	t-ratio	p-value
<b>1. On-Target Comment Frequency: Study Group Predictor</b>				
Intercept, $\beta_{00}$	1.24	0.12	10.27	<.001
... Multiple Baseline Participation	-0.43	0.25	-1.69	.10
Slope, $\beta_{10}$	0.07	0.02	4.08	<.001
... Multiple Baseline Participation	0.04	0.03	1.29	.21
<b>2. Percentage of On-Target Comments: Study Group Predictor</b>				
Intercept, $\beta_{00}$	86.76	2.07	41.89	<.001
... Multiple Baseline Participation, $\beta_{01}$	-11.48	4.18	-2.75	.011
Slope, $\beta_{10}$	0.85	0.32	2.64	.014
... Multiple Baseline Participation, $\beta_{11}$	1.38	0.62	2.22	.035
<b>3. On-Target Comment Frequency: Coding Reliability Predictors</b>				
Intercept, $\beta_{00}$	1.29	0.09	14.41	<.001
... Multiple Baseline Participation, $\beta_{01}$	-0.63	0.18	-3.43	.002
... Behavior Coding Accuracy, $\beta_{02}$	1.99	0.42	4.69	<.001
... Comment Coding Accuracy, $\beta_{03}$	-0.86	0.39	-2.21	.036
Slope, $\beta_{10}$	0.07	0.02	3.94	<.001
... Multiple Baseline Participation, $\beta_{11}$	0.05	0.04	1.48	.15
... Behavior Coding Accuracy, $\beta_{12}$	-0.09	0.08	-1.15	.26
... Comment Coding Accuracy, $\beta_{13}$	0.11	0.07	1.49	.15
<b>4. Percentage of On-Target Comments: Coding Reliability Predictors</b>				
Intercept, $\beta_{00}$	87.38	1.85	47.26	<.001
... Multiple Baseline Participation, $\beta_{01}$	-14.47	3.72	-3.89	<.001
... Behavior Coding Accuracy, $\beta_{02}$	27.51	8.73	3.15	.004
... Comment Coding Accuracy, $\beta_{03}$	-14.72	8.00	-1.84	.078
Slope, $\beta_{10}$	0.78	0.32	2.45	.022
... Multiple Baseline Participation, $\beta_{11}$	1.57	0.62	2.53	.018
... Behavior Coding Accuracy, $\beta_{12}$	-1.42	1.42	-1.00	.33
... Comment Coding Accuracy, $\beta_{13}$	2.83	1.33	2.13	.044
<b>5. On-Target Comment Frequency: Alliance and Consultation Dosage Predictors*</b>				
Intercept, $\beta_{00}$	1.44	0.45	3.18	.004
... Multiple Baseline Participation, $\beta_{01}$	-0.50	0.30	-1.66	.11
... Working Alliance (Clinician), $\beta_{02}$	0.23	0.23	1.01	.33
... Working Alliance (Consultant), $\beta_{03}$	0.48	0.20	2.37	.027
... Total # Consultation Sessions, $\beta_{04}$	-0.01	0.01	-0.50	.63
... # Consultation Prior to ABC Start, $\beta_{05}$	-0.02	0.05	-0.36	.72
Slope, $\beta_{10}$	0.06	0.08	0.81	.43

Effect	Coefficient	SE	t-ratio	p-value
... Multiple Baseline Participation, $\beta_{11}$	0.13	0.05	2.59	.017
... Working Alliance (Clinician), $\beta_{12}$	-0.04	0.04	-1.11	.28
... Working Alliance (Consultant), $\beta_{13}$	0.06	0.03	1.74	.10
... Total # Consultation Sessions, $\beta_{14}$	-0.00	0.00	-1.03	.32
... # Consultation Prior to ABC Start, $\beta_{15}$	0.02	0.01	2.64	.015
<b>6. Percentage of On-Target Comments: Alliance and Consultation Dosage Predictors*</b>				
Intercept, $\beta_{00}$	89.77	8.74	10.27	<.001
... Multiple Baseline Participation, $\beta_{01}$	-9.90	5.77	-1.72	.10
... Working Alliance (Clinician), $\beta_{02}$	1.09	4.38	0.25	.81
... Working Alliance (Consultant), $\beta_{03}$	6.39	3.88	1.65	.11
... Total # Consultation Sessions, $\beta_{04}$	-0.28	0.27	-1.04	.31
... # Consultation Prior to ABC Start, $\beta_{05}$	1.53	1.03	1.48	.15
Slope, $\beta_{10}$	3.31	1.47	2.25	.035
... Multiple Baseline Participation, $\beta_{11}$	1.07	0.98	1.10	.28
... Working Alliance (Clinician), $\beta_{12}$	-0.61	0.74	-0.83	.42
... Working Alliance (Consultant), $\beta_{13}$	-0.42	0.64	-0.66	.52
... Total # Consultation Sessions, $\beta_{14}$	-0.06	0.04	-1.38	.18
... # Consultation Prior to ABC Start, $\beta_{15}$	-0.06	0.18	-0.31	.76
<b>7. On-Target Comment Frequency: Combined Model with Significant Predictors*</b>				
Intercept, $\beta_{00}$	1.20	0.10	12.54	<.001
... Multiple Baseline Participation, $\beta_{01}$	-0.52	0.20	-2.61	.016
... Behavior Coding Accuracy, $\beta_{02}$	1.53	0.45	3.38	.003
... Comment Coding Accuracy, $\beta_{03}$	-0.63	0.39	-1.61	.12
... Working Alliance (Consultant), $\beta_{04}$	0.36	0.16	2.30	.031
Slope, $\beta_{10}$	0.03	0.03	0.94	.36
... Multiple Baseline Participation, $\beta_{11}$	0.09	0.04	1.90	.07
... # Consultation Prior to ABC Start, $\beta_{12}$	0.01	0.01	1.82	.08

*Note.* Fidelity outcomes were coded by consultants. Slope represents a time variable, coded as months (continuous) since beginning to implement ABC. Behavior and comment coding accuracy were the ICCs from clinicians' first 2.5 months of ABC implementation, averaged across behavior and comment codes. Coding accuracy variables and working alliance scores, reported after the conclusion of consultation, were centered around the group mean. SE = Standard Error. For starred models, N = 28 due to missing data for consultant report of working alliance. There is no combined model for percentage of on-target comments because the only significant predictors were coding accuracy variables.



**Table 4**

## Predictors of Linear Growth in Coding Accuracy over Time

Effect	Coefficient	SE	t-ratio	p-value
<b>Behavior Coding Accuracy</b>				
Intercept, $\beta_{00}$	0.65	0.15	4.21	<.001
... Working Alliance (Clinician), $\beta_{01}$	-0.20	0.07	-2.99	.007
... Working Alliance (Consultant), $\beta_{02}$	0.08	0.06	1.38	.18
... Total # Consultation Sessions, $\beta_{03}$	-0.00	0.00	-0.33	.75
... # Consultation Prior to ABC Start, $\beta_{04}$	0.01	0.02	0.33	.74
Slope, $\beta_{10}$	0.09	0.07	1.22	.23
... Working Alliance (Clinician), $\beta_{11}$	-0.07	0.03	-2.04	.053
... Working Alliance (Consultant), $\beta_{12}$	-0.01	0.03	-0.48	.64
... Total # Consultation Sessions, $\beta_{13}$	-0.00	0.00	-0.63	.54
... # Consultation Prior to ABC Start, $\beta_{14}$	0.01	0.01	0.33	.74
<b>Comment Coding Accuracy</b>				
Intercept, $\beta_{00}$	0.71	0.25	2.82	.01
... Working Alliance (Clinician), $\beta_{01}$	-0.04	0.11	-0.32	.75
... Working Alliance (Consultant), $\beta_{02}$	0.07	0.10	0.70	.49
... Total # Consultation Sessions, $\beta_{03}$	-0.01	0.01	-0.84	.41
... # Consultation Prior to ABC Start, $\beta_{04}$	0.02	0.03	0.78	.44
Slope, $\beta_{10}$	0.05	0.10	0.53	.60
... Working Alliance (Clinician), $\beta_{11}$	0.01	0.04	0.20	.84
... Working Alliance (Consultant), $\beta_{12}$	0.01	0.04	0.25	.81
... Total # Consultation Sessions, $\beta_{13}$	-0.00	0.00	-0.15	.89
... # Consultation Prior to ABC Start, $\beta_{14}$	0.00	0.01	0.40	.69

*Note.* Behavior and comment coding accuracy outcomes were ICCs representing clinicians' agreement with consultant coding, averaged across several behavior and comment codes. Time was modeled across four 2.5-month periods, with ICCs representing coding accuracy calculated separately for each period. The model intercepts represent estimated coding accuracy at the conclusion of training. Working alliance scores, reported after the conclusion of consultation, were centered around the group mean. SE = Standard Error. N = 28 due to missing data for consultant report of working alliance.