EFFECTS AND PROCESSES

OF FIDELITY-FOCUSED CONSULTATION

by

EB Caron

A dissertation submitted to the Faculty of the University of Delaware in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Summer 2017

EFFECTS AND PROCESSES

OF FIDELITY-FOCUSED CONSULTATION

by

EB Caron

Approved: _____
Robert F. Simons, Ph.D.
Chair of the Department of Psychological and Brain Sciences

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate & Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Mary Dozier, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Rinad S. Beidas, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Ryan M. Beveridge, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Lisa M. Jaremka, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to acknowledge my family for supporting and encouraging me in the pursuit of a doctoral degree. I especially thank my husband, Dan, for his skilled management of so many aspects of our life together, which allowed me to focus on this dissertation. I would also like to thank my advisor, Mary, for her mentorship, and for guiding me in discovering research interests about which I have become passionate. Finally, I would like to thank my lab sisters, especially Teresa, for their practical help, emotional support, and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Identification of effective consultation models and strategies could inform efforts to implement evidence-based treatments in community settings. Study 1 examined the effects of a fidelity-focused consultation model, using a multiple baseline design with clinicians implementing the Attachment and Biobehavioral Catch-up (ABC) intervention. In fidelity-focused consultation, clinicians received feedback on consultant-rated fidelity coding, and also coded their own fidelity. Clinicians' fidelity to ABC increased in the period immediately after fidelity-focused consultation began, but did not increase during other periods of training. Study 2 used micro-analytic coding of consultation session videos to identify active ingredients within the multi-component package of fidelity-focused consultation. Certain types of fidelity feedback and goal setting predicted future fidelity. Results also suggested that clinicians' self-coding was a critical component of fidelity-focused consultation. Findings from the present study contribute to the literature on evidence-based components of consultation, and suggest that consultation procedures that incorporate fidelity coding feedback and self-coding of fidelity are effective in promoting implementation outcomes.

Chapter 1

STUDY 1

Introduction

Usual care in community mental health settings frequently fails to improve

children's functioning (e.g., Warren, Nelson, Mondragon, Baldwin, & Burlingame,

2010). Although many interventions have proven effective in clinical trials (Chorpita

et al., 2011), usual care clinicians often do not use these interventions (National

Advisory Mental Health Council, 2001; President's New Freedom Commission on

Mental Health, 2004). Frequently, evidence-based interventions do not perform as

successfully in the community as they did in the lab, a discrepancy that is often

attributed to treatment fidelity being lower in community settings than in labs

(Hulleman & Cordray, 2009). Further, community clinicians may believe they are

implementing an evidence-based intervention, but actually exhibit low levels of

intervention fidelity (Santa Ana et al., 2008). The field of implementation science,

which studies the activities that can incorporate a practice into a community agency

(Fixsen, Naoom, Blase, Friedman & Wallace, 2005), has grown in response to these

challenges.

An implementation framework can provide the structure necessary to

understand and respond to various implementation challenges (Tabak, Khoong,

Chambers & Brownson, 2012). Many such frameworks exist (Tabak et al., 2012), but

Fixsen et al.'s (2005) implementation framework is ideal for the current study. Fixsen

et al. (2005) proposed that core intervention practices are taught to clinicians through

communication link processes, within a sphere of influence defined by organizational

structure and culture. Communication link processes include training and coaching,

and fidelity is measured to provide a feedback loop between clinicians and the

communication processes being used to support them. Accurately identifying the core

intervention components of a program is considered critical, as this knowledge defines

the focus for training, coaching, and fidelity evaluation. Fixsen et al. (2005) also

describe stages of implementation, beginning with exploration of possibilities and the

decision to adopt a program, and ending with long-term sustainability, with stages in

between that include initial implementation and full operation.

Across the various stages of implementation, it is critical to identify the active

ingredients of implementation, or the processes that lead to clinician and agency

behavior change. Fixsen et al. (2005) call these active ingredients "implementation

drivers" or core components of implementation. Fixsen et al. (2005) identify staff

selection, training, ongoing consultation/coaching, staff and program evaluation,

facilitative administrative support, and systems interventions as core components of

implementation.

The current study focuses on one of these core components, consultation, with

an emphasis on fidelity feedback loops that inform consultation. Although initial

training workshops often lead to increases in clinicians' knowledge about evidence-

based practices, they rarely change clinicians' behavior (Edmunds, Beidas & Kendall, 2013; Herschell. Kolko, Baumann & Davis, 2010). Consultation or supervision[1] that begins after training and continues while clinicians begin implementing a practice is considered critical in helping clinicians to change their behavior (Edmunds et al., 2013a). Several randomized controlled trials of training and consultation provide support for this idea. Training plus consultation, as compared with a training workshop only, improved clinicians' implementation of motivational interviewing (Schwalbe, Oh & Zweben, 2014), cognitive-behavioral techniques in Multisystemic Therapy (Henggeler, Sheidow, Cunningham, Donohue & Ford, 2008; Holth, Torsheim, Sheidow, Ogden & Henggeler, 2011), and measurement-based care (Lyon, Pullman, Whitaker, Ludwig, Wasse & McCauley, 2017). Web-based training plus consultation was more effective in improving teachers' implementation of evidence-based curriculums than website access alone (Pianta, Mashburn, Downer, Hamre & Justice, 2008). In addition, Beidas, Edmunds, Marcus and Kendall (2012) found a dose-response relationship between consultation and clinicians' implementation of a cognitive-behavioral treatment for anxiety; for every additional hour of consultation

---

[1] Some have conceptualized supervision as internal to a clinician's organization/agency, and consultation as a process with an external expert (Schoenwald, Sheidow & Chapman, 2009; Schoenwald, Sheidow & Letourneau, 2004), while others distinguish consultants from supervisors based on whether they assume direct hierarchical responsibility for patient care (Edmunds et al., 2013a). I try to consistently use the term *consultation* in the present manuscript, but given that both terms, as well as others, have been used variably in the literature (Edmunds et al., 2013a), at times I have instead followed the terms used in original sources.

attended by clinicians, treatment adherence improve by .4 points, and skill improved by .3 points.

However, not all studies of consultation have shown such strong support for its role as an active ingredient of implementation. Although Sholomskas et al. (2005) found a training workshop plus consultation condition superior to providing clinicians with a therapy manual, the consultation condition was not found superior to a web-based training condition in helping clinicians to learn cognitive-behavioral therapy. Similarly, Moyers et al. (2008) did not find the addition of motivational interviewing consultation superior to workshop training alone. In an observational design in which group consultation and individual written feedback were provided, Dunn et al. (2016) failed to find that clinicians' motivational interviewing fidelity increased over time.

These discrepant findings suggest that all forms of consultation are not equal, an idea supported by studies comparing different types of consultation. For example, in a roll-out experimental design, Funderburk et al. (2015) compared standard telephone consultation with live video coaching among clinicians learning Parent-Child Interaction Therapy. They found that the amount of video-based consultation that clinicians had received was associated with greater behavioral improvement in children; however, phone consultation did not predict children's outcomes. Similarly, differential focus on certain activities within the same consultation protocol has been linked to outcomes for clinicians and clients. For example, supervisors' adherence to supervision structure was associated with greater improvement for adolescents receiving Multisystemic Therapy, while greater supervisory focus on clinician

development was associated with poorer improvement for adolescents (Schoenwald, Sheidow, & Chapman, 2009). These findings raise questions about the active ingredients of consultation and the specific processes that make consultation effective (Nadeem, Gleacher & Beidas, 2013).

One potential active ingredient of consultation is provision of feedback. Consultation frequently includes feedback; 63% of the supervisory interventions reviewed by Milne, Aylott, Fitzpatrick and Ellis (2008) included feedback. Contextualized Feedback Intervention Theory (CFIT; Riemer, Rosof-Williams & Bickman, 2005) proposes that feedback in the context of consultation works by alerting clinicians who are committed to a goal that they have not accomplished the goal. Feedback thus creates cognitive dissonance, leading clinicians to change their behavior (Riemer et al., 2005).

Several studies have compared consultation that included feedback with other types of consultation. For example, Noell et al. (2005) examined a consultation strategy for teachers who sought help with student behaviors, that included graphing teachers' implementation of recommended interventions, graphing student behavior, providing positive feedback on correct implementation, and identifying omitted steps. This performance feedback consultation condition was compared to weekly supportive follow-up sessions, with and without an emphasis on commitment to the intervention (Noell et al., 2005). Compared to the other types of consultation, performance feedback consultation led to higher teacher implementation of interventions, as well as greater student behavior change (Noell et al., 2005). Similarly, in an additive within-

subjects design, providing clinicians with graphical feedback on their implementation of a contingency management intervention improved implementation from 42% to 71% (Andrzejewski, Kirby, Morral & Iguchi, 2001).

Feedback based on fidelity coding is seen as the gold standard for feedback in consultation (Dunn et al., 2016). From the perspective of Contextualized Feedback Intervention Theory, using a standardized feedback system like fidelity coding may increase the perceived validity and reliability of feedback, increasing clinicians' attention to and likelihood of accepting such feedback (Riemer et al., 2005). Several studies have examined fidelity feedback consultation. Miller, Yahne, Moyers, Martinez and Pirritano (2004) found that providing written feedback based on a fidelity measure for three audiotaped sessions was as effective as six consultation sessions in helping to maintain clinicians' motivational interviewing training workshop gains at 4 and 8 months post-training. However, only a condition that combined written feedback and consultation led to improvements in client behavior (Miller et al., 2004). Martino et al. (2016) compared a motivational interviewing supervision approach that included performance feedback based on fidelity ratings with supervision as usual, and found greater increases in competence for clinicians receiving fidelity feedback. In addition to these studies of clinicians learning motivational interviewing, Weck et al. (2017) found that, compared with regular consultation, consultation plus written feedback using the Cognitive Therapy Scale led to larger increases in fidelity to cognitive behavioral therapy over time.

Self-monitoring or self-supervisory interventions offer clinicians another form of feedback. Schunk's (1989) self-regulation theory describes a process for self-monitoring that closely parallels the processes proposed by Contextualized Feedback Intervention Theory. In self-regulation theory, self-observation leads to awareness about one's performance (Schunk, 1989). Self-observation is followed by self-evaluation, in which one's performance is compared with one's goals, leading to performance-goal discrepancies and motivation to change behavior (Schunk, 1989).

However, similar to the range of effects observed for consultation, studies of self-supervision have shown an array of outcomes, likely related to how self-supervision is structured and designed. Simple self-monitoring procedures, such as tracking on-schedule and on-task behavior in half hour blocks, have been shown to increase the monitored behaviors, even with minimal training (Richman, Riordan, Reiss, Pyles & Bailey, 1988). More complicated self-supervision procedures, such as reviewing one's taped counseling sessions and recording instances of metaphor and empathy use, have had mixed outcomes with regard to changing clinician behavior, leading the authors to conclude that clinicians may require more training in self-monitoring than was provided (Dennin & Ellis, 2003).

One way to provide such training is to combine self-ratings with peer or expert ratings. Garcia, James, Bischof and Baroffio (2017) designed a faculty development approach for problem-based learning tutors, in which tutors engaged in self-rating using an observational measure immediately after their class, and then again after watching a video of the class. Tutors then met with three peer observers, who had also

watched and rated the video, for a feedback session. Although self-observation did not lead tutors to reappraise their performance, as video-based ratings were unchanged compared to post-class ratings, and tutors rated themselves more positively than their peers, most tutors reported that self-observation was more helpful than peer feedback. In terms of outcomes, tutors previously rated by students as "needing improvement" improved their student ratings the following year, while tutors rated as good maintained these ratings (Garcia et al., 2017).

Though Garcia et al. (2017) created an observational measure specifically for their feedback intervention, Isenhart et al. (2014) adapted an existing motivational interviewing fidelity measure for their group-based consultation protocol. They found that among clinicians who coded their own and others' sessions in groups led by an expert consultant, fidelity to motivational interviewing increased over time. Isenhart et al. (2014) attributed these findings not only to the feedback component of consultation, but also to the process of learning to code fidelity. Specifically, they hypothesized that the process of coding sessions would sensitize clinicians to specific aspects of motivational interviewing fidelity, allow them to quickly and accurately process session content, and provide opportunities to discuss and practice alternative, high fidelity responses. These skills developed in consultation sessions were thought to generalize to clinicians' behavior in their own sessions (Isenhart et al., 2014).

To summarize, understanding how to effectively implement evidence-based interventions in community settings is a critical target for the field. Consultation is considered a core component of implementation, but has not always been found

effective in promoting clinicians' implementation with fidelity. Consultation that includes provision of feedback, particularly fidelity feedback, has shown promise in enhancing implementation outcomes. Self-supervision, with or without concurrent feedback from peers or experts, has been less often utilized as a component of implementation, and represents an area for further exploration.

Attachment and Biobehavioral Catch-up

Attachment and Biobehavioral Catch-up (ABC) is a coaching intervention for parents of infants and toddlers who have experienced early adversity. It is focused on improving the parental behaviors of nurturance (i.e., providing warm, sensitive care in response to children's distress) and following the lead (i.e., responding contingently to children's signals in play and conversation), and reducing parental frightening behavior (i.e., threatening or harsh voice tone or actions). Not only has the intervention been successful in changing parent behavior in both lab (Bernard, Simons & Dozier, 2015) and community settings (Caron, Weston-Lee, Haggerty & Dozier, 2016; Roben, Dozier, Caron & Bernard, in press), but it has also led to a number of positive outcomes for children. Specifically, compared with Developmental Education for Families (DEF), a control intervention focused on physical and cognitive development, ABC improved rates of secure and organized attachment (Bernard et al., 2012). ABC also improved children's diurnal regulation of the stress hormone cortisol, as compared with children in the control intervention (Bernard, Dozier, Bick & Gordon, 2015), effects that were sustained for several years of follow-

up (Bernard, Hostinar & Dozier, 2015). Other effects of ABC observed years after the intervention include improved executive functioning, compared with children assigned to DEF (Lewis-Morrarty, Dozier, Bernard, Terraciano & Moore, 2012; Lind, Raby, Caron, Roben & Dozier, 2017), and differential emotion expression (Lind, Bernard, Ross & Dozier, 2014).

Fixsen et al. (2005) asserted that identifying the core components of a program is critical to inform implementation. The core component of ABC is thought to be "in the moment commenting," a process in which clinicians provide live feedback to parents about their intervention-relevant behaviors, as they interact with their children during sessions. For example, a clinician might say, "That's exactly what we've been talking about with nurturance. He whimpered and looked up at you, and you picked him up and started patting his back." Links to intervention outcomes validate in the moment commenting as a core component of ABC. Specifically, the frequency and quality of clinicians' comments have been found to predict parent behavior change (Caron, Bernard & Dozier, 2016). Comment frequency has also been found to reduce likelihood of early dropout from ABC (Caron et al., 2016a). Following recommendations by Fixsen et al. (2005), the fidelity measure for ABC is focused on measuring its core component, in the moment commenting.

After developing the ABC fidelity measure, we asked a lab-based clinician to self-monitor using the measure, coding one of her ABC sessions each week (Meade, Dozier & Bernard, 2014). This created a single subject A/A+B design, in which the clinician had received standard lab-based group supervision for 6 months, and then

continued standard supervision, with the addition of self-monitoring, for another 6 months. We found that prior to the introduction of self-monitoring, the clinician's comment frequency was unchanging; after beginning self-monitoring, however, her comment frequency increased over time (Meade et al., 2014).

These results encouraged us to develop consultation procedures that incorporated self-monitoring as well as expert feedback using the ABC fidelity measure. The present study examined the effects of this new consultation procedure in a multiple baseline design, in which clinicians first received consultation as usual, and then began receiving both consultation as usual and fidelity-focused consultation. Guided by Fixsen et al.'s (2005) stages of implementation, we continued monitoring clinicians' fidelity for up to 30 months following cessation of consultation, in order to examine the sustainment phase.

Method

Participants

Clinicians

Participants included seven clinicians at different agencies in Hawaii, who were selected for ABC training as part of a non-profit-led implementation effort. Ten clinicians attended the initial training workshop, but one moved to a different position in her agency before beginning to implement ABC. Two other clinicians began implementing ABC with families, but dropped out of training before beginning

11

fidelity-focused consultation. The sample in the current study included the seven

clinicians who participated in fidelity-focused consultation as part of their ABC

training.

All clinicians were female, and all had masters' degrees. Three (43%)

clinicians were European American, one (14%) was Asian American, one (14%) was

Hawaiian American, and two (29%) were multi-racial, with backgrounds including

Native American, Asian, and Hawaiian descent. On average, clinicians were 38 years

old ($SD = 8$ years), and had worked in their jobs for an average of 3.5 years ($SD = 2.4$

years) and in their fields for an average of 10 years ($SD = 5.8$).

Fidelity consultants

Fidelity consultants included seven undergraduate students at the University of

Delaware. All students were female. One student consulted with two clinicians, and

one clinician received consultation from two different students, with a transfer

occurring partway through the consultation period. Additional demographic data were

not collected from consultants.

Procedure

Clinicians were trained in a three-day training workshop in Hawaii in March

2012. After this training, they immediately began group clinical consultation

("consultation as usual"). Clinicians began seeing ABC cases at different times, with

three clinicians beginning to implement ABC in March 2012, two beginning ABC in

April 2012, one in May 2012, and one in June 2012. However, clinicians were expected to attend group consultation regardless of whether they had begun implementing ABC.

Fidelity-focused consultation was intended to begin after clinicians had completed session 4 of ABC with two different families. Due to differences in case load, families dropping out prior to session 4, and clinicians' speed of sending videos to the University of Delaware, clinicians had completed various numbers of sessions (range: 10-27, $M = 20$, $SD = 6$) prior to beginning fidelity-focused consultation. Clinicians began fidelity-focused consultation between June 2012 and September 2012, with two clinicians beginning in June, two clinicians beginning in August, and three clinicians beginning in September. On average, clinicians had been implementing ABC for 3.8 months (SD = 1.2; range: 1.9 – 5.4 months) when they began fidelity-focused consultation. The variability in timing of both beginning to implement ABC and beginning fidelity-focused consultation created an ideal multiple-baseline design.

Clinicians then engaged in fidelity-focused consultation for 10.3 months, on average ($SD = 2.2$; range: 7.7 – 14.3 months). Fidelity-focused consultation was designed to conclude at the end of the training period, when clinicians were certified in ABC. One clinician ended fidelity-focused consultation early (due to maternity leave) in April 2013, three clinicians ended fidelity-focused consultation as planned in May 2013, and three clinicians continued fidelity-focused consultation until August or September 2013. Of the clinicians that continued consultation longer than planned,

two had significant breaks in practicing ABC during the fidelity-focused consultation period (3.7 months and 8.6 months), and the third needed additional time and support to reach certification criteria. During the fidelity-focused consultation period, clinicians completed 74 sessions of ABC, on average ($SD = 41$; range: 32-156), and participated in 27 fidelity-focused consultation sessions ($SD = 9$; range: 11-39).

All clinicians but one were certified in ABC and continued practicing after the training period ended. One clinician, who also had an 8.6-month break in practicing ABC during the training year, decided not to pursue certification for personal reasons and did not continue implementing ABC. Clinicians were asked to send sessions to the University of Delaware for continuing program evaluation, and participation in sending videos varied, with four clinicians sending nearly all of their sessions, and two clinicians sending fewer videos. Two years after certification, clinicians were required to send 10 sessions from a recent case in order to be evaluated for re-certification. All but one of the six certified clinicians submitted sessions for recertification and were recertified. In summary, certified clinicians varied in the number of ABC sessions they conducted after the training period ended, the number of sessions they sent to the University of Delaware, and the timing/spacing of the sessions they sent. Excluding one clinician who was an outlier (386 sessions), on average, there were 35 observations ($SD = 26$; range: 7 – 64) per clinician during the follow-up period. On average, clinicians' final follow-up data point occurred 20 months ($SD = 11$; range: 2 – 31) after consultation ended.

Session videos were coded for the purposes of fidelity-focused consultation and program evaluation, and once coded, archived data lacked identifying information. As such, the University of Delaware Institutional Review Board determined the research exempt and approved archival data analysis.

Consultation Conditions

Consultation as usual

Small group clinical consultation, or consultation as usual, was led by a Ph.D. level consultant, and occurred for one hour each week. It included review of recent session videos, feedback on implementing manual content, and discussion of case-specific dynamics and strategies (e.g., adjusting the approach for a mother who was distracted from interacting with her baby by her active toddler). Consultation as usual also included discussion and feedback on in the moment commenting. However, the clinical consultant did not receive training on ABC fidelity coding until September 2012, and this training was brief and focused on understanding rather than reaching reliability.

Fidelity-focused consultation

Weekly fidelity-focused consultation was implemented by undergraduate students who were reliable ABC fidelity coders. In preparation for fidelity-focused consultation, clinicians were assigned coding from one to three recent session videos. Clinicians sent their coding to their fidelity-focused consultants, who then sent

consultant-rated coding in return. After exchanging coding, clinicians and consultants met for half hour sessions in which clinicians received feedback on both their fidelity and their coding accuracy. Feedback on coding was focused on refining clinicians' understanding of both targeted parent behaviors relevant to the intervention and commenting fidelity (e.g., what counts as a comment, what makes comments off-target). Feedback on fidelity consisted of explicit performance feedback using the fidelity coding, using comparisons to certification criteria and the clinician's prior performance, celebrating gains and strong performance, and providing recommendations and coaching for how to improve performance. For example, for a clinician who struggled with making frequent comments, feedback and coaching might include identification of reasons why (e.g., video clip included clinician playing video feedback to parents on her laptop) and strategies to comment more frequently during the same situation in the future (e.g., "It's okay to pause the video, or play it a few times. You can also keep your comments short when you have your laptop out."). For a clinician who struggled with comment quality (i.e., off-target or low number of components), feedback/coaching would likely include review of individual comments, explanation of why the comments were rated as low-quality, and how to improve them (e.g., "This comment was off-target because you spoke to the child instead of the parent, and we want to make sure the parent is paying attention to what you're saying. So instead of 'Mama picks you up right away when you want a hug,' how do you think you could phrase it differently next time?"). Consultants also used other active learning strategies such as modeling comments and engaging clinicians in role-play of

comments. Fidelity-focused consultants met weekly with the author for group supervision, which included review of fidelity coding and identification of primary areas for feedback to clinicians.

Both consultation as usual and fidelity-focused consultation were conducted remotely using videoconferencing software. Following fidelity-focused consultation sessions, fidelity consultants wrote brief emails about the focus of the session and their primary feedback to the clinical consultant; these emails included fidelity coding, which was archived and used in the current study.

Measure: ABC fidelity

The ABC fidelity measure is focused on a core component of ABC, in the moment commenting. The measure assesses the frequency and quality of clinicians' in the moment feedback to parents. The measure is coded from five-minute video clips of sessions, is reliable, and predicts client outcomes (Caron et al., 2016a).

The fidelity measure involves first coding parent behaviors relevant to the ABC intervention, and then coding clinicians' response (or lack of response) to each behavior. When a clinician fails to respond to a relevant parent behavior, it is coded as a missed opportunity to comment. When a clinician makes an in-the-moment comment, it is coded as either "on-" or "off-target," reflecting whether or not the statement is an appropriate and helpful response to the prior parent behavior. Off-target comments may be ill-timed (e.g., provide praise immediately after a negative parent behavior), provide incorrect information (e.g., tell the parent their behavior was

an example of "following the lead" when it was in fact a "nurturing" behavior), or highlight behaviors that are not targeted by ABC (e.g., provide praise for the parent's monitoring their child without actively responding to him or her). In addition to assessing comment quality through whether comments are on-target, the number of specific information components included in comments (ranging from 0 to 3) are also coded. Information components can include: (1) specifically describing the parent-child interaction ("He rolled you the ball and you rolled it back to him"), (2) labeling the ABC target with which the parent's behavior corresponds ("Beautiful following his lead"), and (3) discussing a long-term outcome the behavior can have on the child ("That's helping him learn he can have an impact on his world."). Comments scored as including 0 information components may still be on-target; typically, such comments provide praise (e.g., "Yes, that's it!") or non-specific descriptions of parent behavior ("Nice responding to him.").

Commenting fidelity is coded on a Microsoft Excel spreadsheet, which calculates summary statistics for the 5-minute clip. In previous work, frequency of on-target comments and percentage of on-target comments (number of on-target comments divided by total number of comments) were found to predict parent behavior change at both the clinician- and case-levels (Caron et al., 2016a), and were chosen as the primary outcomes of interest. Percentage of missed opportunities (parent behaviors not responded to with a comment, divided by total parent behavioral opportunities to comment) and average number of components included in comments

were found to predict parent behavior change at the clinician-level, only, and were examined here as secondary outcomes.

Data from 1217 sessions were available. Excluding one outlier (562 sessions), each clinician had data from an average of 109 sessions ($SD = 42$, range: 52 – 160). One-hundred seventeen videos were double-coded to assess reliability. One-way, single measures, random effects intraclass correlations demonstrated excellent reliability for on-target comment frequency (.95), percentage of on-target comments (.76), and percentage of missed opportunities (.76), and good reliability for average components (.74; Cicchetti & Sparrow, 1981).

Analyses

Piecewise longitudinal hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) was used to estimate separate slopes (i.e., rates of change) for different periods of the implementation process: consultation as usual, fidelity-focused consultation, and sustainment. We split the fidelity-focused consultation period into two five-month periods, a decision guided by analyses in Meade et al. (2014), which split the ABC training period into two 6-month periods, and introduced clinician self-coding in the second 6-month period. Although Meade et al. (2014) found increasing fidelity during the 6 months after the clinician began self-coding, continuing fidelity during a longer period was not assessed. For this reason, we hypothesized that change would occur during the first 5 months of fidelity-focused consultation, but we did not form specific hypotheses about fidelity change during the second 5 months. During

19

the second 5 months after fidelity-focused consultation was introduced, fidelity could continue increasing indefinitely, or could plateau. During the sustainment period, we hypothesized that fidelity would no longer increase, but also would not decline (i.e., slope would not be significantly different from 0). The strongest test of the effects of fidelity-focused consultation would compare growth in fidelity during fidelity-focused consultation to the period before fidelity-focused consultation was introduced (consultation as usual). To test these different hypotheses, we conducted hypothesis tests comparing slopes of different pieces.

We recoded time data for three clinicians to provide accurate slope estimates. Specifically, the 3- and 8-month breaks of two clinicians during the fidelity-focused consultation period were recoded into 1-month breaks, to make these clinicians' training timelines comparable to other clinicians, and because we did not expect to observe improvement during the long breaks in which clinicians were not implementing ABC or engaged in fidelity-focused consultation. Additionally, one clinician who had difficulty meeting certification criteria received 4.5 more months of fidelity-focused consultation than any other clinician. As including this data as-is could decrease the reliability of estimates of change for the group, this clinician's time data were recoded to fit all of the data from her final 9.5 months of fidelity-focused consultation into the second 5-month period (i.e., time data were divided by around 2). Time data representing her first five months of fidelity-focused consultation were left as is. However, changing this final clinician's time data did not change the results;

when run with the original time coding, all results, including hypothesis tests, remained the same.

Piecewise linear growth models were specified estimating fidelity slopes across different time periods in the implementation process (Raudenbush & Bryk, 2002). Models for each fidelity outcome (e.g., percentage of on-target comments, frequency of on-target comments) included four slope variables, which were allowed to vary (e.g., increase) during the period of interest (e.g., consultation as usual) but which were held constant during other periods. Thus, each of the four variables estimated the expected rate of change in fidelity during one of the following periods of interest: consultation as usual, months 1-5 of fidelity-focused consultation, months 6-10 of fidelity-focused consultation, and sustainment.

In addition to allowing this piecewise modeling of change across time, HLM offered a flexible approach to a dataset with significant variability in the number and spacing of time points (Raudenbush & Bryk, 2002). The HLM approach also accounted for the nesting of time-varying observations (session-level fidelity data, level 1) within clinicians (level 2). Nesting at the clinician level accounted for the non-independence of observations within clinicians, that is, for differences in clinicians' initial fidelity and trajectories of change, around the group's average intercept and estimated change. Thus, terms accounting for random effects variance at the clinician level were included in the model for each of the four time periods and at the intercept.

Results

Clinicians demonstrated increasing performance in both primary fidelity outcome measures, comment frequency and percentage of on-target comments, during the first 5 months of fidelity-focused consultation, but not during any other periods. Specifically, as shown in Figure 1, over the course of the first 5 months of fidelity-focused consultation, on average, clinicians showed an estimated increase of 0.77 comments per minute, and 17.8% on-target comments. As shown in Table 1, during the first 5 months of fidelity-focused consultation, slopes for both variables were significantly different from the second 5 months of fidelity-focused consultation and the sustainment period, which had generally flat slopes. Additionally, clinicians showed a larger increase in comment frequency during the first 5 months of fidelity-focused consultation than during the consultation as usual period.

Results for the percentage of missed opportunities, a secondary outcome, mimicked those seen for the primary outcomes. As shown in Table 1 and Figure 2, over the course of the first 5 months of fidelity-focused consultation, clinicians showed decreasing missed opportunities to comment. This result parallels that found for comment frequency, demonstrating that clinicians were commenting more frequently and less frequently failing to respond to relevant parent behaviors. Clinicians did not demonstrate significant change in missed opportunities during any other training periods.

For average number of components included in comments, none of the slopes from any time period were significantly different from 0. However, the rate of change

of the consultation as usual period approached significance ($p = .064$) and hypothesis tests revealed this slope was significantly steeper than the sustainment period. Overall, clinicians appeared to make fairly steady gains across all three training periods, before plateauing in the sustainment period. However, when data were recoded into a 2-piece model (training vs. sustainment), the slope of the training period, $\beta_{10} = 0.04$, still only approached significance ($p = .065$). This failure to find a slope reflecting a consistent pattern of growth among clinicians is likely related to a high amount of variance in the slopes of all three training periods prior to sustainment. As shown in Table 2, significant variance between clinicians was observed in more of the slopes for average number of components, compared with other outcomes.

Thus, for three of four fidelity outcomes, the periods of largest growth occurred during the first 5 months of fidelity-focused consultation. In contrast, for these three outcomes, during the consultation as usual training period, minimal change was observed. There was also minimal change during the second 5 months of fidelity-focused consultation and the sustainment period. The latter is notable, suggesting that clinicians were able to consolidate and maintain gains in fidelity across up to 30 months without additional consultation.

Table 1   Hierarchical Linear Models Testing Commenting Variables as Predictors of Change

Effect	Coefficient	SE	t-ratio	p-value
Comment Frequency (Number of On-Target Comments per Minute)				
Intercept, β_{00}	0.66	0.14	4.81	.003
Consultation as usual, β_{10}	0.02	0.05	0.40	.71
Fidelity consultation months 1-5, β_{20}	0.15	0.03	5.14	.002
Fidelity consultation months 6-10, β_{30}	-0.01	0.03	-0.37	.72
Sustainment, β_{40}	0.00	0.01	0.25	.81
Hypothesis tests: $\beta_{20} > \beta_{10}$ ($p < .05$), $\beta_{20} > \beta_{30}$ ($p < .001$), $\beta_{20} > \beta_{40}$ ($p < .001$)				
Percentage of On-Target Comments				
Intercept, β_{00}	71.15	5.67	12.55	<.001
Consultation as usual, β_{10}	-1.08	2.93	-0.37	.73
Fidelity consultation months 1-5, β_{20}	3.56	1.09	3.26	.017
Fidelity consultation months 6-10, β_{30}	0.35	1.01	0.35	.74
Sustainment, β_{40}	-0.05	0.15	-0.31	.76
Hypothesis tests: $\beta_{20} > \beta_{30}$ ($p < .05$), $\beta_{20} > \beta_{40}$ ($p < .01$)				
Average Number of Comment Components				
Intercept, β_{00}	1.05	0.11	9.74	<.001
Consultation as usual, β_{10}	0.07	0.03	2.27	.064
Fidelity consultation months 1-5, β_{20}	0.04	0.03	1.14	.30
Fidelity consultation months 6-10, β_{30}	0.06	0.04	1.61	.16
Sustainment, β_{40}	-0.01	0.00	-1.71	.14
Hypothesis tests: $\beta_{10} > \beta_{40}$ ($p < .05$)				
Percentage of Missed Opportunities				
Intercept, β_{00}	68.11	3.97	17.17	<.001
Consultation as usual, β_{10}	-0.91	1.45	-0.62	.56
Fidelity consultation months 1-5, β_{20}	-2.88	0.76	-3.77	.009
Fidelity consultation months 6-10, β_{30}	-0.64	1.12	-0.57	.59
Sustainment, β_{40}	0.00	0.20	0.02	.98
Hypothesis tests: $\beta_{20} > \beta_{40}$ ($p < .01$)				

Note. The estimates in the coefficient column represent the monthly expected change in the fidelity outcome.

24

Table 2   Estimation of Random Effects Variance Components

Effect	SD	Variance Component	χ^2	*p*-value
Comment Frequency (Number of On-Target Comments per Minute)				
Intercept, r_{00}	0.29	0.083	13.11	.022
Consultation as usual, r_{10}	0.05	0.002	2.81	>.50
Fidelity consultation months 1-5, r_{20}	0.04	0.002	4.94	>.50
Fidelity consultation months 6-10, r_{30}	0.01	0.000	3.38	>.50
Sustainment, r_{40}	0.02	0.000	11.63	.040
Percentage of On-Target Comments				
Intercept, r_{00}	13.21	174.38	23.99	<.001
Consultation as usual, r_{10}	6.70	44.94	27.41	<.001
Fidelity consultation months 1-5, r_{20}	1.99	3.95	8.35	.14
Fidelity consultation months 6-10, r_{30}	1.61	2.60	7.80	.17
Sustainment, r_{40}	0.08	0.01	1.72	>.50
Average Number of Comment Components				
Intercept, r_{00}	0.26	0.066	36.17	<.001
Consultation as usual, r_{10}	0.05	0.002	14.01	.015
Fidelity consultation months 1-5, r_{20}	0.08	0.007	18.00	.003
Fidelity consultation months 6-10, r_{30}	0.09	0.007	35.59	<.001
Sustainment, r_{40}	0.01	0.000	6.43	.27
Percentage of Missed Opportunities				
Intercept, r_{00}	8.99	80.87	12.92	.02
Consultation as usual, r_{10}	2.59	6.73	9.12	.10
Fidelity consultation months 1-5, r_{20}	1.18	1.39	4.65	>.50
Fidelity consultation months 6-10, r_{30}	2.25	5.07	8.45	.13
Sustainment, r_{40}	0.37	0.14	7.94	.16

Note. Chi-squares are based on 6 of 7 clinicians that had sufficient data for

computation; degrees of freedom are 5.

Comments Per Minute
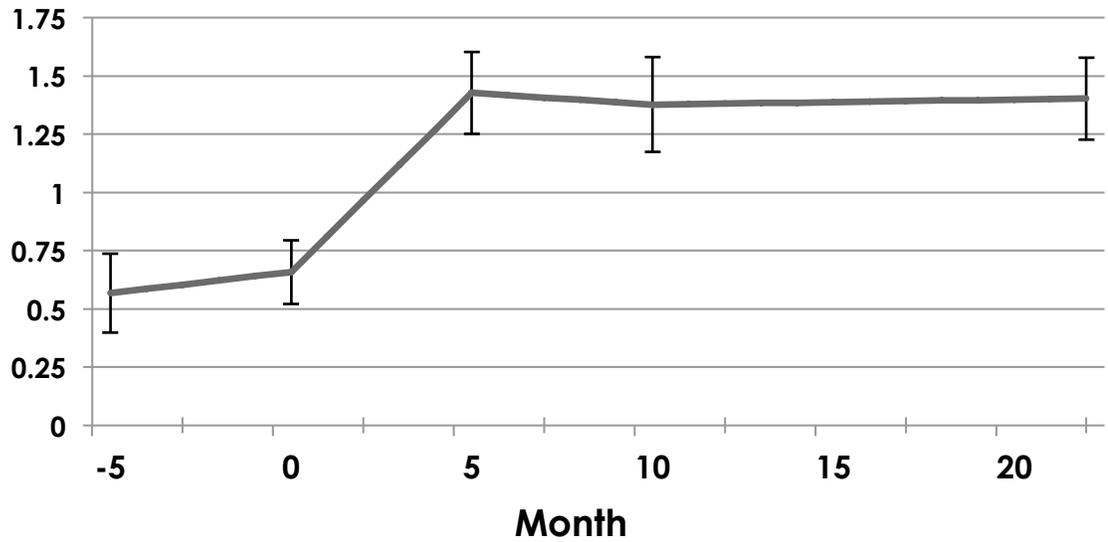
Percentage of On-Target Comments
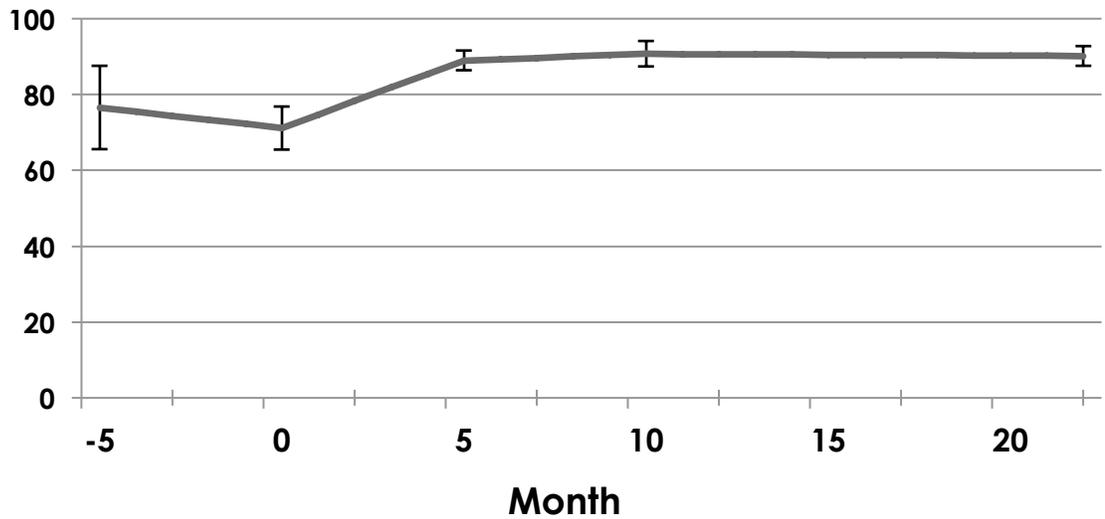
Figure 1   Primary outcomes. Graphs show only the first year of the sustainment period. However, data were modeled through up to 31 months of sustainment (which would be Month 41 on the current graphs).

Average Comment Components
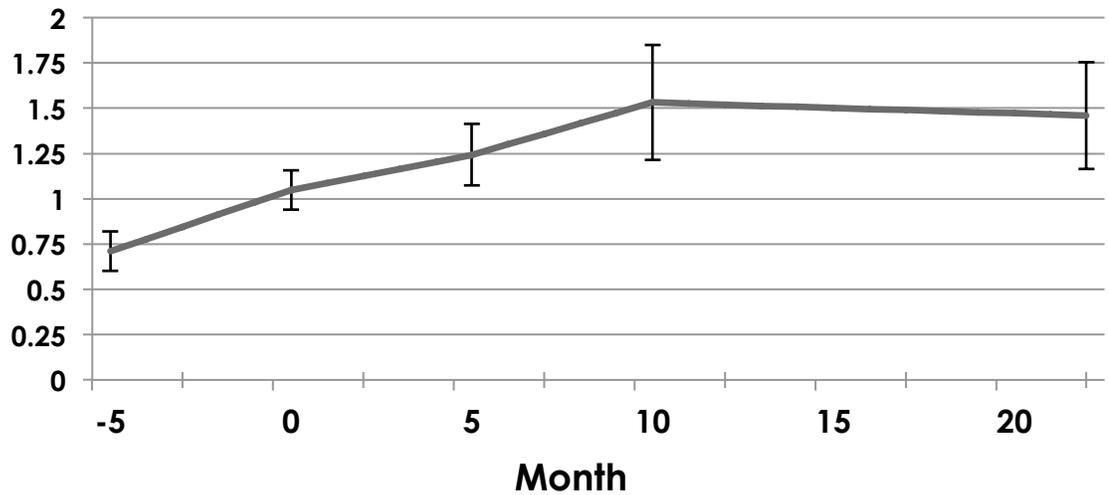
Percentage of Missed Opportunities
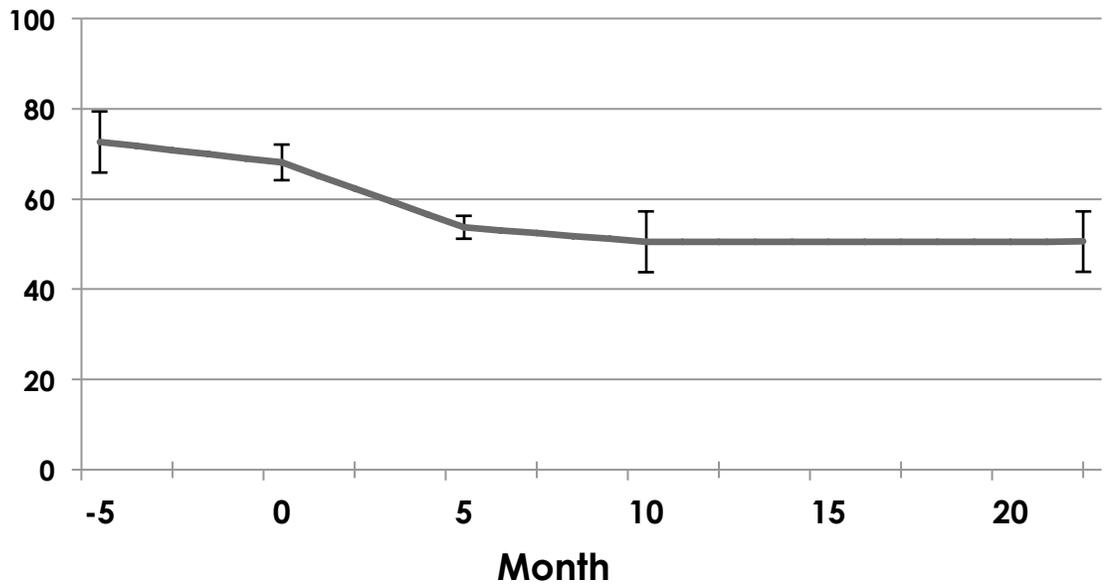
Figure 2   Secondary outcomes. Graphs show only the first year of the sustainment period. However, data were modeled through up to 31 months of sustainment (which would be Month 41 on the current graphs).

Discussion

The current study examined the impact of fidelity-focused consultation, in which clinicians engaged in self-coding and received fidelity feedback from an expert coder on a weekly basis. The results suggest that this consultation procedure increased clinicians' growth in ABC fidelity over time. These results are consistent with other studies that have found that performance feedback using fidelity data can result in continued improvement following training (Martino et al., 2016; Weck et al., 2017) or can prevent post-training skill decline (Miller et al., 2004; Noell et al., 2005). It is notable that these results were found when fidelity-focused consultation was added to an already intensive and evidence-supported consultation protocol, weekly hour-long consultation using video review of recent sessions. That is, while the gold standard of clinical consultation is still being elaborated (Nadeem et al., 2013), Garland and Schoenwald (2013) identified supervision using video or audio review as a quality control method that was effective in promoting treatment fidelity. Further, the intensity of the consultation as usual condition in the current study stands in contrast to some studies, such as Martino et al. (2016), in which only 5 supervision as usual sessions occurred during the trial, as compared with 247 sessions of the performance feedback supervision.

A unique aspect to the performance feedback in the current study, as compared with most prior work, is that clinicians were engaged in coding their own fidelity. This component of the consultation procedure may have increased clinicians' receptivity to consultants' feedback, by increasing perceived source credibility and

feedback validity, which may have increased the impact of the feedback intervention (Riemer et al., 2005). Thus, the self-supervision component of fidelity-focused consultation may have contributed to its strong impact on fidelity, which was observed above and beyond the effects of consultation as usual. However, this hypothesis remains to be tested, and could be tested in a trial in which clinicians are randomized to two conditions of fidelity-focused consultation, one similar to the current study, and one in which clinicians receive fidelity coding feedback but do not engage in self-coding.

To some, the findings of the current study may not be impressive; the process of fidelity coding feedback leading to changes in fidelity may seem like "teaching to the test." Indeed, after Dunn et al. (2016) found that their less structured, lower cost feedback intervention did not increase motivational interviewing fidelity, they concluded, "It may be necessary to 'teach to the test.' In other words, despite the cost of giving providers ongoing feedback about their [fidelity] scores, perhaps that it what is needed to raise [fidelity] scores over time" (p. 81). In contrast, Noell et al. (2005) arguably taught to the test with their successful performance feedback intervention; consultation sessions with teachers included reviewing data to score implementation, graphing implementation, providing positive feedback about completed intervention components, and identifying components that were omitted or implemented incorrectly. Consultation sessions occurred daily until teachers implemented the intervention with 100% fidelity, then occurred every other day until teachers achieved 2 consecutive days with 100% fidelity, and then occurred weekly (Noell et al., 2005).

Although "teaching to the test" has a negative connotation, I believe that this undertone does not apply to fidelity feedback interventions, and particularly to this study, for two reasons. First, the "test" has been linked to desired outcomes for families (Caron et al., 2016a), showing its value and importance. Second, clinicians maintained performance on the "test" for over two years after they have stopped receiving "grades," showing that what they learned was maintained far beyond the period of testing.

In fact, this notable lack of decline in fidelity across 30 months of follow up is one of the study's biggest strengths. Skill levels following training and consultation have been shown to be maintained for 6 months in studies of clinicians learning Multisystemic Therapy (Henggeler et al., 2008a) and measurement-based care (Lyon et al., 2017). Post-consultation levels of fidelity have further been maintained for 12 to 15 months among clinicians learning motivational interviewing (Persson et al., 2016) and Trauma-Focused Cognitive Behavioral Therapy (Murray, 2017). The over two year sustainment period in the current study, particularly with its demonstrated lack of deterioration in fidelity, is rare in the literature, and speaks to the effectiveness of ABC consultation procedures.

Another finding of the current study was the lack of growth in fidelity during the second 5 months of fidelity-focused consultation. The study that led to the development of fidelity-focused consultation (Meade et al. 2014) examined only 6 months of self-coding, and at end of this period, found fidelity increasing with a quadratic slope. The current study suggests that such growth may not have continued

with additional time spent self-coding. However, the feedback intervention in the current study was also more intensive than that in Meade et al. (2014), with clinicians receiving, on average, 15 sessions of fidelity-focused consultation ($SD = 3$, range: 10 to 18) in the first 5 months, with feedback on an average of 27 sessions of ABC ($SD = 7$, range: 19 to 40) during that time. In comparison, in Meade et al. (2014), the clinician coded 15 video clips during the 6 months of self-coding, and did not receive feedback on her coding or fidelity.

The lack of improvement in fidelity during the second 5 months of fidelity-focused consultation leads to questions about the necessary dose of such consultation. The results of the current study suggest that clinicians may not need a full 10 months of consultation to achieve the levels of ABC fidelity required for certification in ABC. However, it is also possible that the second 5 months of fidelity-focused consultation are necessary to achieve the sustainment of fidelity observed during the extended follow-up period. Applying Fixsen et al.'s (2005) stages of implementation, the first 5 months of fidelity-focused consultation likely correspond to the "awkward" initial implementation stage, in which skills levels are changing, while the second 5 months may correspond to the full operation stage, which occurs when new learning is integrated into clinicians' practice, and clinicians implement a program with competence. It may be important for clinicians to experience the full operation stage while still receiving expert consultation. Specifically, during the second 5 months of fidelity-focused consultation, consultants may engage clinicians in processes that consolidate previously learned skills and teach self-supervisory practices that

31

clinicians can use after consultation ends. Consultants may also focus on advanced

commenting techniques, such as commenting on parents' negative behaviors, which

were not examined as markers of fidelity in the current study. Future work should

investigate the dose of fidelity-focused consultation necessary to both achieve *and*

sustain ABC certification requirements over an extended follow-up period. Future

research can also examine moderators of the necessary dose of fidelity-focused

consultation; for example, clinicians who begin with weaker skills or who intervene

with fewer clients may need more time to reach ABC certification requirements.

Indeed, the results presented here are group patterns; as described in the Method

section, one of the clinicians in the current study required an additional four months of

consultation to meet certification requirements. Future work could also investigate the

impact of thinning the frequency of fidelity-focused consultation sessions after ABC

certification requirements have been met; for example, decreasing consultation

frequency to biweekly sessions in the second half of training (Noell et al., 2005).

Similar to questions about the necessary dose of fidelity-focused consultation

are questions about the importance and necessary dose of consultation as usual.

Because consultation as usual was a constant throughout the training period in the

current study, it is possible that fidelity-focused consultation only leads to increased

fidelity when clinicians are receiving more traditional consultation concurrently.

Additionally, due to the current study's multiple-baseline design, it is also possible

that order/timing effects may be present, and the rapid response to fidelity-focused

consultation is only observed after clinicians have received traditional consultation for

several months. Future work could examine relative impacts of traditional consultation, fidelity-focused consultation, and their combination in a group design. If traditional consultation is found to be important in fidelity outcomes, additional work could explore different doses of traditional consultation, by varying both frequency and duration.

Questions about necessary doses of fidelity-focused and traditional consultation are important because the more cost effective that ABC training can be, the more easily and widely it can be implemented (Herschell et al., 2010; Weisz, Ugueto, Herren, Afienko & Rutt, 2011). Two aspects of fidelity-focused consultation lower its cost. First, only five-minute video clips are coded; typically, this coding can be completed in 30 to 60 minutes. Second, all of the coder-consultants in the current study were undergraduate coders, who received course credit as compensation for their time. Although part-time undergraduates cannot supervise as many clinicians as full-time staff, requiring a greater time investment in training consultants, the model in the current study shows that advanced degrees are not required to provide high-quality, effective fidelity-focused consultation. In fact, the clinicians in the current study rated their satisfaction with consultation and their working alliance with consultants equally high for fidelity-focused and traditional Ph.D.-level consultants (Caron et al., 2016b).

A limitation of the current results is that only one of four fidelity outcomes, comment frequency, showed statistically stronger performance during fidelity-focused consultation compared with consultation as usual. Although the rate of growth in

percentage of on-target comments during the first 5 months of fidelity-focused

consultation, 3.56% per month, was significantly different than 0, it was not

significantly different than the consultation as usual period, which was -1.08% per

month (i.e., a non-significant negative growth rate). The failure to find these numbers

different may relate to the small sample size at level 2 and a significant amount of

unexplained variance in the consultation as usual slope (see Table 2), reflecting

notable between-clinician variance in the slope during that period. Additionally, the

consultation as usual period was shorter and had a smaller number of observations per

clinician than other periods of observation, which may have lowered power to detect

differences between the slope of this period and other periods. However, the other

two fidelity variables that did not show slope differences between the fidelity-focused

consultation period and the consultation as usual period had been identified as

secondary outcome variables, as they have shown less impact on parent behavior

change.

A major weakness of the current study is the small sample size at level 2,

which limited statistical power. Hand in hand with the small sample size was the lack

of randomization to intervention conditions, which would have further reduced power

by splitting the sample into smaller groups. Because there was no control group,

making this an observational design, the argument for causality is weakened. We

cannot conclude that fidelity-focused consultation caused improvements in fidelity; it

is possible that the observed changes were part of the typical training trajectory of

clinicians learning ABC, independent of fidelity-focused consultation.

A strength of the current study was the number of observations per clinician, which allowed for nuanced modeling of fidelity across time, and reliable estimates of clinicians' behavior. In a motivational interviewing study, the number of observations needed per clinician for a reliable estimate of fidelity with real patients ranged from 5 to 13, depending on the marker of fidelity (Imel et al., 2014). The high number of observations per clinician may be even more important with the fidelity measure used in the current study, in which only 5 minutes of sessions are coded, as short segments may not accurately represent full sessions, but do reasonably reflect performance when aggregated (Friedlander et al., 1988).

This study was designed to test the impact of fidelity coding feedback and self-coding as a consultation intervention. Although the intervention as a package appears effective, it is critical to understand why. That is, although fidelity coding feedback is hypothesized to be the primary active ingredient of fidelity-focused consultation, this theory remains to be tested. Other components of fidelity-focused consultation, such as modeling or role-play, which have been shown effective in other studies of consultation (Bearman et al., 2013; Edmunds et al., 2013b), may in fact be driving the results of the current study. Thus, to further validate the theory and process of fidelity-focused consultation, both fidelity coding feedback and other potential active ingredients of consultation must be measured and tested as predictors of change in clinicians' fidelity. For this reason, in the next study, I will examine different components of fidelity-focused consultation, and assess their links to change in fidelity during ABC training.

Chapter 2

STUDY 2

Introduction

Understanding implementation processes is critical to overcoming the gap between lab-based clinical treatment trials and the psychotherapeutic treatments available in communities (Fixsen et al., 2005). Consultation is one such implementation process. However, it is not well understood (Nadeem et al., 2013). Although consultation is generally regarded as a process that promotes implementation with fidelity (Edmunds et al., 2013a), in certain studies, consultation has not been effective, and the reasons for these findings have been unclear (e.g., Dunn et al., 2016). Other studies have shown that certain types of consultation are more effective than others (e.g., Funderburk et al., 2015), but there is little consensus across studies on the ingredients, processes, and dosages that make consultation effective (Edmunds et al., 2013a; Nadeem et al., 2013; Weisz et al., 2011). As such, empirical testing of the impact of consultation strategies and consultation models on implementation outcomes is needed (Nadeem et al., 2013).

The first study tested the impact of a model of fidelity-focused consultation on clinicians' growth over time in fidelity to the ABC intervention. In a multiple baseline design in which clinicians began implementing ABC while receiving group clinical consultation as usual, growth in fidelity was strongest in the period immediately after fidelity-focused consultation was introduced. Although the consultation model as a

package appeared effective, it was unclear why it had the effects it did. Though the model was designed to be a feedback intervention, it included several other potential active ingredients, including self-evaluation, client outcome assessment and discussion, and active learning techniques such as modeling and role-play. This mix of potential active ingredients in a model of consultation is typical; in fact, Milne et al.'s (2008) review of supervisory interventions identified 26 different components of supervision, with the average study including 5.4 components.

<p align="center">Feedback</p>

Feedback was the second most common component of supervisory interventions reviewed by Milne et al. (2008), observed in 63% of studies. However, meta-analytic results suggest that over one-third of feedback interventions have negative effects on performance (Kluger & DeNisi, 1996). Two theories of feedback processes, Feedback Intervention Theory (Kluger & DeNisi, 1996) and Contextualized Feedback Intervention Theory (Riemer et al., 2005), explain these findings by elaborating processes that may moderate the effects of feedback. First, the two theories assert that behavior is regulated through comparisons between feedback and goals or standards. Discrepancies create cognitive dissonance, which can motivate behavior change (Riemer et al., 2005). Behavioral change processes are regulated by limited attention, so by directing the focus of attention with specific feedback, feedback interventions can change behavior (Kluger & DeNisi, 1996). Behavior change is only one possible response to feedback, however; other possible outcomes

include changing one's goal, becoming less committed to one's goal, attributing one's performance to external factors, and rejecting the feedback (Kluger & DeNisi, 1996; Riemer et al., 2005).

Various aspects of feedback can moderate its effects. Feedback Intervention Theory proposes a hierarchy of behavior change processes, ranging from low-level task-learning processes, to mid-level task-motivation processes, to higher-level self-goal processes. Highly specific feedback that directs attention to lower-level task-learning processes is most likely to lead to behavior change (Kluger & DeNisi, 1996). In contrast, feedback that directs attention to the self and activates higher-level self-goal processes, through comparison of performance with norms or even by providing praise, is thought to worsen performance by directing cognitive resources away from the task (Kluger & DeNisi, 1996). Feedback that makes comparisons with prior performance is thought to activate motivational processes and improve performance (Kluger & DeNisi, 1996). When feedback makes comparisons of performance with "ought" standards, attention is focused on prevention goals, leading to efforts to meet task standards; however, if the "ought" standard is exceeded, feedback may lead to declines in performance (Kluger & DeNisi, 1998). Finally, when goals are clear, feedback is more likely to change behavior (Kluger & DeNisi, 1996). Goal setting is thought to clarify goals and direct attention to lower-level task processes, improving the effects of feedback interventions (Kluger & DeNisi, 1996). Goal setting also creates opportunities for feedback that compares performance with previously set

goals, which is thought to active task-learning processes and increase the effectiveness of feedback (Kluger & DeNisi, 1998).

Self-Directed Feedback

Feedback can also be self-driven. Schunk's (1989) theory of self-regulated learning closely parallels feedback theories. Schunk (1989) describes stages of self-observation, self-judgment and self-reaction. In self-observation, individuals assess their behavior; Schunk (1989) notes that self-recording tends to promote the accuracy of self-observation. In self-judgment, individuals compare their performance with their goals; the specificity of and standards referenced in goals moderate self-judgment processes. Finally, in self-reaction, individuals respond to their self-judgments by taking behavioral, affective or cognitive action, for example, increasing their motivation, or rewarding themselves for their performance.

Engagement in structured self-monitoring is relatively rare in consultation procedures; Milne et al. (2008) noted it as a supervisory technique in only one (4%) study reviewed. Dennin and Ellis (2003) examined a stand-alone self-supervisory intervention that involved viewing and coding one's therapy videos, and found mixed effects in clinical skill development for trainees. They noted that trainees rated themselves more highly than independent observers, and tended to set vague, nonspecific goals for themselves, which may have attenuated the effects of self-monitoring. Dennin and Ellis (2003) concluded that additional training in these self-supervisory skills may have enhanced intervention effects. In that vein, Garcia et al.

39

(2017) and Isenhart et al. (2014) paired self-monitoring with provision of expert ratings and feedback, and observed growth in monitored skills.

Like Dennin and Ellis (2003), Garcia et al. (2017) noted that teachers tended to rate their performance higher than observers, joining a literature that has historically suggested that clinicians are not reliable raters of, and tend to overestimate, their own therapeutic performance (Carroll, Nich & Rounsaville, 1998; Hurlburt, Garland, Nguyen & Brookman-Frazee, 2010; Martino, Ball, Nich, Frankforter & Carroll, 2009). In contrast, other recent studies have provided some optimism about clinicians' ability to evaluate their own performance (Hogue, Dauber, Lichvar, Bobek & Henderson, 2015; Peavy et al., 2014). Further, increasing the amount of training that clinicians receive in rating is likely to increase clinician-observer reliability (Carroll et al., 1998; Hogue et al., 2013; Hurlburt et al., 2010; Peavy et al., 2014). To my knowledge, only one study has examined change in clinicians' ability to self-rate over time. Loades and Myles (2016) examined Cognitive Therapy trainees' self-ratings on the Cognitive Therapy Scale – Revised (CTS-R), as compared with expert raters, over time. Although limited by a small sample size (n = 13), they observed a trend toward lower average discrepancies with expert raters over time. Although this change occurred while clinicians were completing a 30-week training course, it is unclear whether training on CTS-R coding was a component of the course. Thus, more research is needed on the effects of training clinicians in self-rating fidelity, as clinicians' overestimation of their performance is likely to lead to inaccurate self-judgments, reducing the effectiveness of self-feedback processes.

Client Outcome Feedback

Client outcome monitoring provides another type of feedback that has been linked to implementation outcomes. In a randomized trial in a community setting, weekly outcome monitoring, conducted through brief client self-report measures, improved client outcomes (Bickman, Kelley, Breda, de Andrade & Riemer, 2011). When used in consultation, client outcome monitoring is theorized to counter clinicians' and consultants' tendencies to overestimate client progress, and to provide independent information (Worthen & Lambert, 2007). It can help consultants and clinicians recognize client problem areas, focus consultation on strategies that will address these problems, and thereby enhance client outcomes (Worthen & Lambert, 2007). Further, client outcome monitoring offers a standardized source of performance feedback to clinicians (Worthen & Lambert, 2007). Riemer et al. (2005) theorized that systematic client outcome feedback could activate clinicians' motivation to change their own behavior. From the perspective of Feedback Intervention Theory (Kluger & DeNisi, 1996), client outcome feedback together with discussion of targeted problem-solving strategies may activate task-learning processes, and thereby change clinicians' behavior. On the other hand, client outcome feedback on its own, or "mere knowledge of results," may interfere with complex task learning (Kluger & DeNisi, 1996; Riemer et al., 2005).

Pairing clinician fidelity data with client outcome data may be a particularly compelling consultation strategy. In Noell et al.'s (2005) performance feedback condition, teachers were provided with graphs of their own implementation fidelity, as

well as student behavior, across time. Additionally, though not measured numerically, Pianta et al. (2008) described using video feedback to show teachers how students responded to teaching behaviors. These feedback strategies may have helped teachers to observe links between intervention implementation and positive student outcomes, increasing their valuing of and commitment to implementation fidelity. The ABC fidelity measure, which requires coding both parent behaviors and clinician responses, enables consultants to use this feedback technique.

Other Active Learning Strategies

In addition to consultants' feedback and goal setting, clinicians' self-directed feedback, and client outcome feedback, several other strategies are potential active ingredients of fidelity-focused consultation. In recent years, active learning strategies, in particular, modeling and role-play, have gained an evidence base. In modeling, consultants demonstrate intervention techniques, while in role-play, clinicians practice intervention techniques. Both are fairly common consultation techniques, with modeling observed in 29%, and role-play in 21%, of Milne et al.'s (2008) reviewed studies of effective supervision. Bearman et al. (2013) found that both role-play and modeling predicted increased clinician implementation of consultant-recommended intervention strategies. Role-play and modeling appeared particularly helpful for older clinicians (Bearman et al., 2013). In a study of group consultation processes, role-play did not have a main effect on clinicians' future fidelity; however, clinicians' involvement in consultation moderated this association, such that clinicians who were

more involved in consultation appeared to benefit from role-play (Edmunds et al., 2013b). Qualitative research has raised concerns about the acceptability of role-play to clinicians, particularly in group consultation contexts, in which role-play may trigger self-consciousness and performance anxiety (Beidas et al., 2013).

Use of video observation is another evidence-based component of consultation. Live or video-recorded observation was observed in 42% of Milne et al.'s (2008) reviewed studies of supervision, and video or audio observation was used in 57% of the supervisory protocols reviewed by Garland and Schoenwald (2013). Video can promote consultants' accurate perception of both client and clinician behavior in sessions (Funderburk et al., 2015), and thereby refine consultants' feedback and recommendations. For clinicians, viewing video of one's performance can lead to higher self-efficacy, motivation, and self-regulated strategy use in subsequent skill practice (Schunk, 1989).

Working Alliance

In addition to these evidence-based components of consultation, non-specific factors may also predict implementation outcomes. It is possible that the Study 1 findings were driven by the additional support and attention provided to clinicians, rather than the content of fidelity-focused consultation. Meta-analytic work on therapeutic alliance between clinicians and clients suggests it has a moderate and consistent association with treatment outcomes (Martin, Garske & Davis, 2000). Clinician-consultant alliance has been infrequently studied as a predictor of

implementation outcomes, with mixed results. One study found that teacher-reported alliance with their consultant predicted teachers' implementation of a preventative behavioral intervention (Wehby, Maggin, Partin & Robertson, 2012). However, a study of Multisystemic Therapy (MST) consultation found that after controlling for other aspects of consultation quality (consultant competence, adherence to MST procedures), higher clinician-reported alliance with the consultant was associated with lower clinician fidelity and increased adolescent externalizing and internalizing problems (Schoenwald, Sheidow & Letourneau, 2004). Thus, a currently small literature on clinician-consultant alliance has produced mixed results, suggesting that additional research is needed in this area.

<center>Dose</center>

Another possible interpretation for the Study 1 findings is that the sheer amount, or dose, of consultation led to improved fidelity. That is, perhaps the doubled dose of consultation (i.e., traditional consultation plus fidelity-focused consultation) led to the gains in fidelity after fidelity-focused consultation was introduced. One meta-analysis found that greater doses of motivational interviewing consultation, in terms of both total hours and longer timeline, predicted clinicians' fidelity (Schwalbe et al., 2014). Similarly, Beidas et al. (2012) found dose-response links between consultation and clinicians' fidelity to a cognitive-behavioral treatment. Funderburk et al. (2015) also found a dose-response association between consultation and client

outcomes. Thus, dose should be considered as an alternative predictor of links between consultation and implementation outcomes.

Organizational Context

Fixsen et al. (2005) assert that implementation components "do not exist in a vacuum"; to capture this, their implementation framework depicts the influence of organizational and external context as a constant background to implementation processes (p. 58). Indeed, organizational context matters; Beidas et al. (2015) found that 7-23% of variation in clinicians' use of therapy techniques was attributed to organizational characteristics, and that for cognitive-behavioral and family therapy, organizational characteristics accounted for greater variance than individual factors. Organizational factors have also been specifically linked to fidelity to evidence-based treatments; for example, organizational readiness to change predicted clinicians' fidelity to cognitive-behavioral components of contingency management (Henggeler et al., 2008a). Nuances in these associations have also been found; for example, organizational climate predicted Multisystemic Therapy fidelity for clinicians with low fidelity, but not for clinicians with high fidelity (Schoenwald, Sheidow, Letourneau & Liao, 2003). Thus, organizational characteristics are not just a background for implementation, but rather a key factor, and should therefore be considered as alternative predictors of implementation outcomes.

Present Study

The present study examined components of fidelity-focused consultation and their role as predictors of change in ABC fidelity. Feedback, goal setting, self-directed feedback, client outcome feedback, active learning techniques, working alliance, dose, and perceived organizational support were considered as potential predictors of implementation outcomes. Videos of fidelity-focused consultation sessions were coded using a micro-analytic coding system of consultant and clinician processes, and these consultation processes were examined as predictors of both next-session fidelity and long-term fidelity trajectories. Clinicians' self-rating reliability was also explored, as both an outcome of fidelity-focused consultation, as well as a potential predictor of change in fidelity.

Method

Participants

Clinicians

Participants included 27 clinicians from 18 different agencies in 6 US states. Three additional clinicians consented to the current study, but were not included in analyses because of insufficient data, due to early termination of ABC training related to maternity leave ($n = 1$) and leaving the agency ($n = 1$), and lack of recorded fidelity-focused supervision sessions ($n = 1$). Full demographic data for 1 of the 27 clinicians were not available because she started but did not complete the questionnaires.

All clinicians but one (96%) were female. Most (n = 18, 67%) clinicians had Masters degrees; 6 (22%) had Bachelors degrees, 1 (4%) had a Ph.D., and 1 (4%) had completed high school or less. Most (n = 16, 59%) clinicians were White, 6 (22%) were Black, 1 (4%) was Asian American, and 3 (11%) were more than one race. One White clinician and one Black clinician (7%) were Hispanic. On average, clinicians were 34 years old (*SD* = 7.5, range: 23 – 48). They reported having worked an average of 4.8 years (*SD* = 5.7) in their current jobs, and 9.1 years (*SD* = 6.7) in the field.

Fidelity consultants

Twenty-one fidelity consultants participated in the current study. All consultants but one (95%) were female. Most fidelity consultants were White/non-Hispanic (n = 13, 62%), 6 (29%) were White/Hispanic, and 1 (5%) was Black. Average reported age was 22 (*SD* = 1.7, range: 20 – 26). The majority (n = 14, 67%) were undergraduate students, and 7 (33%) worked as part- or full-time staff. Of the staff members, 6 had completed their Bachelors degree and 1 had a Ph.D.

Switches in clinician-consultant pairings occurred when 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 (n = 15, 56%) clinicians received consultation from a single consultant during their training period. However, 9 (33%) had 2 consultants during their training, and 3 (11%) had 3 consultants.

Procedure

Clinicians were trained in a two- to three-day training workshop in 2012, 2015 or 2016. For the clinicians trained in 2012, fidelity-focused consultation began several months after they started implementing ABC; full details on their training timeline can be found in Study 1. The rest of the clinicians began both fidelity-focused consultation and general ABC consultation soon after the training workshop and typically before they began implementing ABC.

All clinicians were provided with a session-by-session ABC manual, as well as access to a training website that included videos of ABC sessions conducted by expert clinicians. Prior to 2013, fidelity-focused consultation was closely supervised by the author in weekly group meetings; these weekly group meetings continued throughout the study period. Fidelity-focused consultation procedures were manualized in 2013, and consultation training was formalized into a series of weekly hour-long meetings in 2015. Thus, most clinicians received fidelity-focused consultation guided by a manual, and about half (n = 12) of the consultants in the current study participated in the formalized consultation training meetings.

Fidelity-focused consultation sessions were conducted as described in Study 1. Fidelity coding was archived from consultants' recap emails and used in the current study. Consultation sessions were videotaped for training and feedback for consultants; these videos were archived and used in the current study. The study protocol was approved by the University of Delaware Institutional Review Board (IRB), and clinicians provided informed consent for use of their archived materials as

well as for completing questionnaires about working alliance, agency support, and demographics. Clinicians received a $20 gift card as compensation for completing the questionnaires.

<center>Measures</center>

ABC Fidelity

ABC fidelity was assessed using the measure described in Study 1. For the clinicians who also participated in Study 1, only fidelity data coded for the purposes of consultation were used so that the same method was used across the entire Study 2 sample. Data from 1038 sessions were available; on average, each clinician had data from an average of 38 sessions (SD = 14, range: 22 – 76). Sixty-nine videos were double coded by other consultants. One-way, single measures, random effects intraclass correlations (ICCs) demonstrated excellent reliability for on-target comment frequency (ICC = .84), and good reliability for percentage of on-target comments (ICC = .66), percentage of missed opportunities (ICC = .72), and average components (ICC = .72; Cicchetti & Sparrow, 1981).

Components of Fidelity-Focused Consultation

A system for coding content of fidelity-focused consultation sessions was developed through consultation of the literature (Edmunds et al., 2013b; Goodyear, 2014; Milne et al., 2008; Milne, James, Keegan & Dudley, 2002; Milne, Pilkington, Gracie & James, 2003) and progressive testing of the coding system. Coding was

<center>49</center>

completed in Noldus the Observer XT software, which allows for coding of subjects, behaviors, and modifiers. Subjects included the consultant, the clinician, and other. Consultant behaviors included giving feedback, providing information/recommendations, asking questions or prompting, discussing clients, goal setting, expressing support or understanding, modeling, and live coding (i.e., discussing or explaining coding using a previously uncoded video clip or worksheet). Clinician codes included many of the same behaviors coded for supervisors, including feedback, questions, client discussion, goal setting, and live coding. Unique behaviors coded only for clinicians included reflection (i.e., expressing thoughts or reactions to coding, commenting, or other ABC-related situations), responding to consultant input (i.e., briefer responses that did not express thoughts), and role play. Playing of video without anyone speaking was coded under the "other" subject.

Different modifiers were coded for different behavior codes. Feedback had the most extensive system of modifiers, which included overall topic of feedback (commenting or coding), specific topic of feedback (e.g., comment frequency or behavior coding), valence of feedback (positive, neutral, or negative), reference (norms, past performance, goals, certification criteria, developmental expectations, effect on client, just numbers without comparison, consultant coding, and other) and specificity (single comment/behavior, in 5-minute coded clip, and more general). Other modifiers can be seen in the coding manual, which is in Appendix B.

Six consultation sessions per clinician were coded, with a total of 162 coded sessions. For each clinician, one session per month, during the first 6 months after

beginning to implement ABC, was selected for coding. Only consultation sessions for which fidelity coding was completed were coded. Continuous, second-by-second coding was completed for full consultation sessions, and each variable's percentage of the overall session time was selected as the measurement of interest.

Thirty-two videos (20%) were selected for double coding. Two-way mixed (consistency) ICCs, shown in Table 3, demonstrated excellent reliability for all broad consultant variables, with the exception of live coding (ICC = .64), for which reliability was good (Cicchetti & Sparrow, 1981). As modifiers were added to consultant variables, reliability tended to fall into the good, fair or poor range, although for some variables, reliability remained excellent. Many variables that demonstrated poor reliability occurred rather infrequently in the dataset. As shown in Table 4, reliability was poor for a number of clinician variables, including some broad variables, such as goal setting and self-directed comment feedback. Clinician self-directed coding feedback, client discussion, role-play and live coding had excellent reliability.

Working Alliance

Clinician-consultant working alliance was measured using the Working Alliance Inventory - Trainee and Supervisor Versions (WAI-T and WAI-S; Bahrick, 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. In the current

sample, internal consistency of the WAI-T and WAI-S were good (Cronbach's alpha =

.88 and .91, respectively). Full scale scores were negatively skewed and were inverse-

transformed. Both clinicians and consultants rated alliance at the conclusion of

fidelity-focused consultation for each consultant or clinician they had worked with. In

the current study, when clinicians received consultation from multiple consultants,

scores were averaged to create one WAI-T score and one WAI-S score per clinician.

Perceived Organizational Support

Clinicians' beliefs about support provided by their organization were

measured using the Survey of Perceived Organizational Support – Short Form (SPOS;

Rhoades, Eisenberger & Armeli, 2001). The measure includes 8 items such as "My

organization cares about my opinions." Respondents make ratings on 7-point Likert

scales, with some items reverse-scored. In the current sample, internal reliability for

the scale was acceptable (Cronbach's alpha = .80). Full scale scores were negatively

skewed and were log-transformed.

<center>Analyses</center>

Data Cleaning

To create time variables for longitudinal analyses, ABC session dates were

taken from fidelity coding sheets. However, 44 dates from 6 clinicians were missing.

In these cases, session dates were estimated using the clinician's coding date (entered

<center>52</center>

on another line of the coding sheet), the consultant's coding date, and the consultation session date.

As shown in Tables 3 and 4, many components of consultation occurred relatively infrequently, and as a result, there were many positively skewed variables with a high frequency of 0 values. Most variables were either log-transformed or inverse-transformed, depending on the extent of the skew and which transformation better corrected the skew. Tables 3 and 4 list type of transformation used. Frequency of on-target comments and percentage of on-target comments were also skewed, and were log-transformed. As a result of these transformations, a number of variables were inverted; to facilitate interpretation of significant results, in the written results, signs are corrected to reflect the true direction of effects. All original signs are maintained in tables.

Session-to-Session Models

Data from the 162 coded consultation sessions were paired with the clinician's fidelity data from their next session of ABC. Associations between consultation session data and future fidelity were explored with bivariate correlations. Significant correlations were probed in HLM, with the consultation session variables entered as predictors at level 1, to test whether associations would hold after accounting for the nested structure of the data (i.e., repeated measurements within clinicians). Models took the form of $Fidelity_{ti} = \beta_{00} + \beta_{10}*Predictor_{ti} + r_{0i} + r_{1i}*Predictor_{ti} + e_{ti}$, with β_{10} representing the variable of interest.

Linear Growth Models

Following data cleaning procedures from Study 1, time gaps in ABC implementation that were longer than 1 month were recoded as 1 month gaps. Session dates for 8 clinicians were recoded, and following recoding, the average length of fidelity-focused consultation was 8.9 months ($SD = 1.5$). Because this length of time was similar to the length of fidelity-focused consultation in Study 1, time data were recoded into two 5-month pieces for piecewise analyses.

First, to attempt to replicate findings from Study 1, piecewise linear growth models were specified, estimating fidelity slopes across the first and second 5 months of ABC implementation. Models took the form: $Fidelity_{ti} = \beta_{00} + \beta_{10}*Piece1_{ti} + \beta_{20}*Piece2_{ti} + r_{0i} + r_{1i}*Piece1_{ti} + r_{2i}*Piece2_{ti} + e_{ti}$, with β_{10} and β_{20} representing the slope variables of interest. Next, consultation variables were tested as level 2 predictors of change in fidelity across those periods, adding an error term and either "+ $\beta_{11}*Predictor_i*Piece1_{ti}$" or "+ $\beta_{21}*Predictor_i*Piece2_{ti}$" to the model. For each clinician, data from the 6 consultation sessions were averaged to create summary level predictors. These were entered individually in separate models, predicting either change in fidelity across the first 5 months or the second 5 months of ABC implementation. To limit the number of statistical analyses, only the primary outcome measures, on-target comment frequency and percentage of on-target comments, were used in these analyses.

Alternative predictors, including working alliance, perceived organizational support, and dose, were also tested as level 2 predictors of change in fidelity. In

testing alternative hypotheses, in addition to testing prediction of growth during consultation, predictors were also tested at the intercept, which allowed examination of associations with initial fidelity.

Clinicians' Coding Reliability

Clinicians' coding reliability was examined across the full group using one-way, random effects ICCs. Single measures ICCs were used, in order to compare reliability with our expert sample, in which only some sessions were double-coded. In addition to group ICCs, individuals' coding reliability was also examined using one-way, single measures, random effects ICCs. To examine change in coding reliability across time, data were split into four 2.5-month time periods. Individuals' reliability statistics for a given time period were excluded if they were based on fewer than 5 data points.

Clinicians' growth in coding reliability across time was examined using bivariate correlations, and hierarchical linear growth models with the following form: $Reliability_{ti} = \beta_{00} + \beta_{10}*Time_{ti} + r_{0i} + r_{1i}*Time_{ti} + e_{ti}$. Predictors of change in coding reliability were explored by adding a level 2 predictor to the growth model ("+ $\beta_{11}*Predictor_i*Time_{ti}$"). Additionally, coding reliability during the first 2.5 months and 5 months of consultation were tested as predictors of concurrent and later fidelity, as described above in the linear growth modeling analysis section.

Table 3   Descriptive Statistics and ICCs for Consultant Components of
             Consultation

Consultant Variables	T	Duration (sec.) Mean (SD)	% Time Coded Mean (SD)	% Time Coded ICC
Comment Feedback (CF)	N	152.8 (97.4)	9.64 (5.66)	.80
CF – Negative	I	4.4 (12.6)	.29 (.84)	.74
CF – Neutral	L	80.8 (70.3)	4.99 (4.13)	.45
CF – Positive	L	67.6 (55.7)	4.36 (3.57)	.84
CF – Single Comment	L	87.1 (73.6)	5.51 (4.47)	.62
CF – 5 Minute Clip	L	55.1 (44.3)	3.49 (2.55)	.79
CF – General	I	10.4 (18.6)	.63 (1.12)	.79
CF – Client	I	22.7 (37.0)	1.47 (2.46)	.65
CF – Developmental	--	3.3 (12.2)	.19 (.67)	.08
CF – Goals	I	3.6 (9.4)	.24 (.64)	.92
CF – Numbers (without any comparison)	L	26.0 (28.7)	1.71 (1.91)	.57
CF – Norms/Other Clinicians	I	3.6 (9.8)	.23 (.63)	.48
CF – Past Performance	--	6.2 (13.0)	.42 (.85)	.18
CF – Unclear	L	79.2 (64.1)	4.86 (3.53)	.58
CF – Certification	I	8.8 (15.7)	.57 (.95)	.74
CF – Certification – Positive	I	4.0 (8.9)	.26 (.53)	.51
CF – Certification – Neutral/Neg. – 5 Min	I	4.4 (10.5)	.29 (.69)	.44
CF – Certification – Neutral/Neg. – General	--	.4 (2.1)	.02 (.14)	.00**
CF – Client Reference – Positive	I	11.8 (25.8)	.76 (1.62)	.45
Coding Feedback	L	145.3 (103.6)	9.47 (7.04)	.87
Coding Feedback – Negative	--	3.8 (10.0)	.25 (.74)	.00**
Coding Feedback – Neutral	L	122.0 (94.3)	7.94 (6.33)	.87
Coding Feedback – Positive	I	19.5 (23.9)	1.28 (1.61)	.64
Coding Feedback – Single Behavior	L	111.1 (93.6)	7.28 (6.38)	.80
Coding Feedback – 5 Minute Clip	L	30.5 (26.3)	1.97 (1.66)	.61
Coding Feedback – General	--	3.7 (9.2)	.22 (.53)	.05
Comment Goal Setting	I	8.9 (17.9)	.57 (1.11)	.78
Comment Goal Setting – Most Specific	--	.1 (.7)	.00 (.06)	n/a*
Comment Goal Setting – Moderately Specific	I	5.8 (13.9)	.37 (.89)	.87
Comment Goal Setting – Nonspecific	I	3.1 (7.5)	.20 (.47)	.48

Note. T = Transformation. N = no transformation needed. L = log-transformed. I = inverse transformed. Only variables with reliability > .40 were considered for analyses and transformed.
* Zero variance for both coder 1 and 2 (percent agreement 100%) ** Zero variance for coder 2.

Table 3 continued.

Coding Goal Setting	I	1.6 (13.1)	.07 (.52)	.81
Coding Goal Setting – Most Specific	I	.2 (1.8)	.01 (.09)	.76
Coding Goal Setting – Moderately Specific	--	.1 (.9)	.00 (.05)	-.01
Coding Goal Setting – Nonspecific	--	1.3 (12.9)	.06 (.51)	-.02
Client Discussion	L	39.8 (46.9)	2.54 (3.32)	.83
Client Discussion – Progress	I	9.1 (20.1)	.59 (1.25)	.51
Client Discussion – Link to Fidelity	I	9.8 (17.8)	.64 (1.20)	.51
Client Discussion – Progress & Link to Fidelity	--	2.7 (8.8)	.17 (.55)	-.04
Modeling	I	19.9 (22.1)	1.16 (1.16)	.92
Live Coding	I	11.7 (31.3)	.70 (1.77)	.64

Table 4   Descriptive Statistics and ICCs for Clinician and Other Components of
             Consultation

Clinician Variables	T	Duration (sec.) Mean (SD)	% Time Coded Mean (SD)	% Time Coded ICC
Comment Feedback (CF)	--	12.6 (18.1)	.89 (1.36)	.00
CF – Certification	--	.2 (1.14)	.01 (.07)	n/a*
CF – Negative	--	5.1 (10.2)	.34 (.69)	.06
CF – Neutral	--	6.5 (11.3)	.47 (.90)	.19
CF – Positive	--	1.1 (4.4)	.08 (.35)	.09
CF – Single Comment	--	4.3 (11.3)	.30 (.82)	.09
CF – 5 Minute Clip	--	4.2 (9.2)	.29 (.66)	.07
CF – General	I	4.1 (9.0)	.30 (.69)	.80
CF – Client Reference – Positive	--	.5 (3.5)	.03 (.24)	.00**
Coding Feedback	I	8.6 (15.9)	.56 (1.09)	.85
Coding Feedback – Negative	--	.6 (2.6)	.04 (.15)	.08
Coding Feedback – Neutral	I	7.6 (15.4)	.50 (1.07)	.87
Coding Feedback – Positive	I	.3 (1.6)	.02 (.10)	.99
Coding Feedback – Single Behavior	I	4.6 (11.0)	.30 (.83)	.90
Coding Feedback – 5 Minute Clip	I	3.6 (8.5)	.24 (.60)	.57
Coding Feedback – General	--	.4 (2.0)	.02 (.11)	.26
Comment Goal Setting	--	2.6 (6.6)	.16 (.43)	.16
Comment Goal Setting – Moderately Specific	--	.8 (3.7)	.05 (.26)	.00**
Comment Goal Setting – Nonspecific	--	1.8 (5.3)	.11 (.33)	.16
Coding Goal Setting	--	.2 (1.7)	.02 (.12)	.00***
Coding Goal Setting – Moderately Specific	--	.1 (.9)	.01 (.07)	n/a*
Coding Goal Setting – Nonspecific	--	.1 (1.1)	.01 (.08)	.00***
Client Discussion	L	79.7 (86.9)	4.85 (4.86)	.76
Client Discussion – Progress	I	14.9 (28.6)	.86 (1.57)	.60
Client Discussion – Link to Fidelity	--	16.0 (25.7)	.95 (1.45)	.27
Client Discussion – Progress & Link to Fidelity	--	4.1 (11.0)	.23 (.62)	.23
Role Play	I	7.3 (14.7)	.44 (.88)	.93
Live Coding	I	3.0 (9.0)	.19 (.57)	.75
Other				
Silence while Watching Video	I	28.0 (62.0)	1.69 (3.55)	.99

Note. T = Transformation. L = log-transformed. I = inverse transformed. Only
variables with reliability > .40 were considered for analyses and transformed.
* Zero variance for both coder 1 and 2 (percent agreement 100%) ** Zero variance for
coder 2. *** Zero variance for coder 1.

Hypotheses

1. Feedback as a global construct will not predict growth in fidelity, but features of feedback may moderate the effects of feedback on fidelity. Positively-valenced feedback is unlikely to be associated with fidelity growth, but neutral or negatively-valenced feedback is more likely to motivate behavior change. More specific feedback is also likely to be associated with growth in fidelity, as it is likely to activate task-learning processes. Providing feedback that the clinician has not met certification criteria (an "ought" standard) will predict growth in fidelity, particularly when such feedback is accompanied by specific information to help the clinician improve commenting. Conversely, providing feedback that the clinician has exceeded certification goals will not predict growth in fidelity, and may even predict declines.

2. Goal setting will predict growth in fidelity throughout the training year. Specific goals will be most successful in promoting fidelity growth.

3. Clinicians' self-directed feedback is likely to predict growth in fidelity. Moderators of self-directed feedback may be similar to moderators of consultants' feedback (e.g., feedback that one has exceeded goals or standards is unlikely to predict change in fidelity).

4. Discussion of client progress, particularly as a response to the clinician's use of in the moment feedback, will be associated with increases in clinician fidelity.

5. Active learning strategies, including modeling, role play, live coding, and watching video, will predict fidelity growth.

6. Feedback on coding will lead to increased self-coding reliability, which in turn will be associated with future fidelity. Specific feedback and neutral or negatively-valenced feedback is more likely to lead to growth in coding reliability and fidelity than nonspecific or positively-valenced feedback.

Results

Session-to-Session Models

Hypothesis 1: Fidelity Feedback and Features of Feedback

Consultant provision of fidelity feedback was not associated with frequency of comments or percentage of on-target comments in the clinician's next ABC session, *r* (162) = .02 and -.02, *ns*, respectively. Examining more specific aspects of feedback for which reliability was at least fair (i.e., >.40; Cicchetti & Sparrow, 1981), the valence of feedback (positive, negative, or neutral) and level of specificity of feedback did not predict fidelity in the next session, with nonsignficant *r*'s ranging from -.09 to .12. Feedback that compared the clinician's performance to goals, norms, the client, or certification criteria, or had an unclear or nonspecific reference was not associated with future commenting, with nonsignficant *r*'s ranging from -.03 to .13. However, feedback in which the consultant discussed the raw numbers on the fidelity coding sheet (without comparing them with past performance, certification criteria, etc.) was associated with the clinician making more comments in the following ABC session, *r* (162) = .24, *p* < .01, but not future percentage of on-target comments, *r* (162) = .13,

ns. The association between this type of feedback and future comment frequency was probed in HLM, and held when accounting for the nesting of data, $\beta_{10} = .14$, t (26) = 2.94, $p < .01$.

Based on a priori hypotheses, combinations of specificity and valence were explored for feedback that referenced certification criteria. Feedback that neutrally or negatively evaluated the clinician's performance as compared with certification criteria, with a level of specificity of the 5-minute coded clip, was associated with making more comments in the next session, r (162) = .18, $p < .05$, but not future percentage of on-target comments, r (162) = -.02, *ns*. The association between this type of feedback and future comment frequency held in HLM, $\beta_{10} = .33$, t (26) = 2.14, $p < .05$. Feedback that positively compared the clinician's performance to certification criteria was not associated with future fidelity, r (162) = -.07 and .02, *ns*.

Hypothesis 2: Goal Setting and Goal Specificity

Consultants' setting of commenting goals for clinicians was not associated with clinicians' future fidelity, r (162) = -.07 and .04, *ns*. However, consultants' setting of coding goals was associated with clinicians making fewer comments in the next session, r (162) = -.17, $p < .05$, and a similar association with percentage of on-target comments approached significance, r (162) = -.15, $p = .064$. These associations remained significant or approaching significance in HLM, $\beta_{10} = -.58$, t (26) = -2.22, $p < .05$, and $\beta_{10} = -2.43$, t (26) = -1.91, $p = .068$. The specificity of goals that consultants set for clinicians did not predict future fidelity, with nonsignificant r's ranging from -

.10 to .06. Clinicians' goal setting was not examined as a predictor of future fidelity

due to poor reliability.

Hypothesis 3: Self-Directed Feedback

Clinicians' self-directed comment feedback was not examined as a predictor of

future fidelity due to poor reliability. Clinicians' self-directed coding feedback was

not associated with either fidelity outcome, r (162) = -.01 and .00, ns. Specificity and

valence of coding feedback did not change these nonsignificant associations, with

nonsignificant r's ranging from -.04 to .05.

Hypothesis 4: Discussion of Client Progress and Links to Fidelity

Consultants' discussion of clients, including case conceptualization, progress,

and other topics, was associated with clinicians making fewer comments in the next

ABC session, r (162) = -.19, p < .05, but not with clinicians' percentage of on-target

comments, r (162) = .04, ns. When consultants made links to commenting fidelity in

their discussion of clients, clinicians made fewer comments in their next ABC session,

r (162) = -.17, p < .05, but were not more often off-target, r (162) = -.04, ns.

However, when examined in HLM, neither of these significant correlations held after

accounting for nesting of observations within clinicians, β_{10} = -.06, t (26) = -1.55, p =

.13, and β_{10} = -.18, t (26) = -1.61, p = .12. Consultants' discussion of client progress

was not associated with either outcome, r (162) = -.07 and .12, ns, respectively.

Consultants could also link commenting to client outcomes in positive fidelity

feedback that used the client as a reference; this type of feedback was associated with the clinician having a higher percentage of on-target comments in the next ABC session, $r(162) = .16$, $p < .05$, but was not associated with future comment frequency, $r(162) = .06$, *ns*. Again, this significant association did not hold in HLM, $\beta_{10} = .74$, $t(26) = 1.74$, $p = .09$.

Clinicians' discussion of clients did not predict future fidelity, $r(162) = -.09$ and .10, *ns*. Clinicians' discussion of client progress was also not associated with either future commenting outcome, $r(162) = .00$ and .07, *ns*, respectively.

Hypothesis 5: Active Learning Strategies

Active learning strategies were not associated with future comment frequency or percentage of on-target comments, including use of video, $r(162) = .10$ and -.00, *ns*, consultant modeling, $r(162) = -.03$ and -.06, *ns*, and clinician role-play, $r(162) = .06$ and .02, *ns*. Likewise, neither consultant nor clinician "live coding" were associated with fidelity outcomes, with nonsignificant r's ranging from -.04 to .04.

Hypothesis 6: Coding Feedback

Consultants' coding feedback was not associated with comment frequency or percentage of on-target comments in the following ABC session, $r(162) = .07$ and .10, *ns*, respectively. However, specificity of consultants' coding feedback moderated associations with comment frequency. That is, consultants' feedback about coding at the level of the 5-minute clip was associated with clinicians making more frequent

comments in the following ABC session, r (162) = .19, $p < .05$, but this association

did not hold when controlling for nesting in HLM, $\beta_{10} = .09$, t (26) = 1.70, $p = .10$.

Consultants' coding feedback that was more specific (i.e., regarding a specific

behavior), was not associated with comment frequency, r (162) = -.03, ns. Neutral

and positively valenced coding feedback were also not associated with comment

frequency, r (162) = .06 and .05, ns, respectively. Specificity and valence did not

moderate associations of coding feedback with future percentage of on-target

comments, with nonsignificant r's ranging from .02 to .12.

Longitudinal Models

Change in Fidelity over Time

As shown in Table 5, clinicians demonstrated increasing comment rate and

percentage of on-target comments, and decreasing missed opportunities, over the first

5 months of ABC implementation. As in Study 1, they did not show change in the

average number of components included in comments. Unlike in Study 1, clinicians

demonstrated continued improvement in comment frequency over the second 5

months of ABC implementation, as shown by both increasing comment rate and

decreasing missed opportunities. Comparison of model intercepts suggested that the

Study 2 clinicians started implementing ABC with better fidelity than the Study 1

group. Perhaps because the clinicians in Study 2 started implementation with stronger

fidelity, they appeared not to improve fidelity as rapidly across the first 5 months of

implementation, as suggested by somewhat smaller slope coefficients.

Table 5   Fidelity Growth over Consultation Period

Effect	Coefficient	SE	t-ratio	p-value
Comment Frequency (Number of On-Target Comments per Minute)				
Intercept, β_{00}	1.15	.11	10.01	<.001
Fidelity consultation months 1-5, β_{10}	.08	.03	2.65	.014
Fidelity consultation months 6-10, β_{20}	.07	.03	2.42	.023
Percentage of On-Target Comments				
Intercept, β_{00}	83.16	2.42	34.33	<.001
Fidelity consultation months 1-5, β_{10}	1.58	.56	2.81	.009
Fidelity consultation months 6-10, β_{20}	.50	.54	.92	.37
Average Number of Comment Components				
Intercept, β_{00}	1.38	.07	20.95	<001
Fidelity consultation months 1-5, β_{10}	-.01	.02	-.81	.43
Fidelity consultation months 6-10, β_{20}	-.00	.01	-.25	.81
Percentage of Missed Opportunities				
Intercept, β_{00}	58.20	2.43	23.92	<.001
Fidelity consultation months 1-5, β_{10}	-1.88	.63	-2.98	.006
Fidelity consultation months 6-10, β_{20}	-1.29	.62	-2.09	.047

Note. Coefficients and standard errors in the table are non-transformed, so that values are interpretable and comparable with Study 1.

Hypotheses about Components of Consultation

Table 6 presents *t*-values and significance levels from linear growth models testing associations between components of consultation and change in fidelity across the first and second 5 months of ABC implementation. Many variables that were hypothesized to be associated with growth in fidelity, such as comment feedback that referenced goals or certification criteria, discussion of client progress, modeling, and role-play, were not associated with change in fidelity. Four variables were associated with change in fidelity, about what would be expected by chance given the number of tests run. Consultants' comment goal setting, particularly that which was more specific, was associated with greater improvement in percentage of on-target comments during the first 5 months of ABC implementation, $\beta_{11} = .32$, $t(25) = 2.30$, $p < .05$ and $\beta_{11} = .44$, $t(25) = 2.48$, $p < .05$. Consultants' setting of highly specific coding goals was associated with poorer improvement in frequency of on-target comments during the first 5 months, $\beta_{11} = -.65$, $t(25) = -2.16$, $p < .05$. However, clinicians' self-provision of positively-valenced coding feedback was associated with greater improvement in frequency of on-target comments during the first 5 months, $\beta_{11} = .69$, $t(25) = 3.07$, $p < .01$. Because of the number of tests run, it is likely that some of these associations occurred by chance and may not reflect associations that would generalize to other studies of consultation.

Table 6   T-tests of Consultation Components as Predictors of Linear Growth in Fidelity

Predictor Variable	Comment Rate Piece 1 t	Piece 2 t	Percent On-Target Piece 1 t	Piece 2 t
Consultant Variables				
Comment Feedback (CF)	-.35	-1.60	.59	-.05
CF – Neutral	-.34	-1.78	1.26	.26
CF – Negative	1.16	.83	-.88	-.88
CF – Positive	.56	-.81	-1.02	-.80
CF – Single Comment	-.81	-1.44	.20	-.44
CF – 5 Minute Clip	1.02	-1.13	.24	-.17
CF – General	-.44	.35	1.60	.91
CF – Client	.91	1.38	.13	.39
CF – Goals	-.42	.50	1.00	.23
CF – Numbers (w/o any comparison)	1.22	-1.71	1.69	.08
CF – Norms/Other Clinicians	.14	.25	-1.05	-.82
CF – Unclear	-.12	-.63	-.52	-.42
CF – Certification	-.84	1.11	.22	.66
CF – Certification – Positive	-1.74	-.36	.97	1.00
CF – Certification–Neutral/Neg.–5 Min	.38	1.88	-1.24	-.89
CF – Client Reference – Positive	.56	.95	1.63	.85
Coding Feedback	.57	-.35	-.75	-1.21
Coding Feedback – Single Behavior	.50	-.35	-1.21	-1.52
Coding Feedback – 5 Minute Clip	.27	-.65	1.00	.85
Coding Feedback – Neutral	.85	-.38	-.93	-1.36
Coding Feedback – Positive	-.06	.28	-.32	-.18
Comment Goal Setting	-.21	-1.07	2.30*	1.60
Comment Goal Setting – Moderately Specific	.23	-.91	2.48*	1.42
Comment Goal Setting – Nonspecific	-.40	-1.22	1.82	1.27
Coding Goal Setting	1.07	.04	.43	-.89
Coding Goal Setting – Most Specific	2.16*	.53	-.25	-.82
Client Discussion	-1.40	-.69	-.71	-.23
Client Discussion – Progress	.09	.92	1.37	.64
Client Discussion – Link to Fidelity	1.16	.29	1.32	-.01
Modeling	.96	.72	.01	-.48
Live Coding	-.14	-.66	-.86	-1.32

Note. * *p* < .05; ** *p* < .01

Table 6 continued.

Clinician Variables				
Comment Feedback – General	-.83	-.69	1.18	1.53
Coding Feedback	-.42	-.25	-.14	-.21
Coding Feedback – Single Behavior	-.73	-.54	-.39	-.70
Coding Feedback – 5 Minute Clip	-.29	-.30	.20	.47
Coding Feedback – Neutral	-.07	-.05	-.20	-.30
Coding Feedback – Positive	-3.07**	-1.64	-.06	-.15
Client Discussion	-.54	-.03	.33	-.13
Client Discussion – Progress	-.03	.26	1.27	.39
Role Play	.57	-.36	.13	-.21
Live Coding	-.69	-.67	-.70	-.69
Other Variables				
Silence while Watching Video	1.05	-.58	-.33	-.64

Alternative Predictors

Table 7 presents the *t*-values and significance levels of tests of associations

between alternative predictors and initial fidelity, as well as change in fidelity across

the first and second 5 months of ABC implementation. Two variables predicted

clinicians' initial fidelity (i.e., the model intercept). The number of fidelity-focused

consultation sessions that clinicians attended prior to starting ABC was associated

with stronger initial percentage of on-target comments, $\beta_{01} = .28$, $t\,(25) = 2.35$, $p < .05$,

suggesting that there was a benefit to meeting with a consultant even before starting to

code one's own sessions. Additionally, consultant-reported working alliance was

associated with greater initial comment frequency, $\beta_{01} = .65$, $t\,(24) = 3.12$, $p < .01$.

Consultant-reported alliance trended toward predicting growth in comment frequency

over the first 5 months of consultation, $\beta_{11} = .08$, $t\,(24) = 2.05$, $p = .051$.

Dose, as measured by both the total number of fidelity-focused consultation sessions and the duration of consultation in months, was associated with slower growth in percentage of on-target comments during the first 5 months of consultation, $\beta_{11} = -.01$, $t(25) = -2.79$, $p < .05$ and $\beta_{11} = -.22$, $t(25) = -2.88$, $p < .01$, respectively. These measures of consultation dose were also associated with slower growth in percentage of on-target comments during the second 5 months of consultation, $\beta_{21} = -.01$, $t(25) = -2.53$, $p < .05$ and $\beta_{21} = -.37$, $t(25) = -3.60$, $p < .01$, respectively. Finally, the number of months of consultation predicted lower initial percentage of on-target comments when beginning to implement ABC, $\beta_{01} = -.76$, $t(25) = -2.08$, $p < .05$. Because dosage of consultation could be adjusted to ensure clinicians were able to reach certification criteria, what these associations likely show is that when clinicians struggled with making on-target comments and were slow to improve their ability to make on-target comments, consultation continued longer than for clinicians who struggled less.

Finally, clinicians who reported lower perceived organizational support showed more rapid growth in frequency of comments during both the first and second 5 months of consultation than clinicians who reported higher organizational support, $\beta_{11} = -.04$, $t(24) = -2.30$, $p < .05$ and $\beta_{21} = -.05$, $t(24) = -2.54$, $p < .05$. Clinician-reported working alliance, the average length of consultation sessions in minutes, and the frequency of consultation sessions were not associated with initial ABC fidelity or growth in fidelity over time.

Table 7   T-tests of Alternative Predictors of Linear Growth in Fidelity

Alternative Predictors	Comment Rate *t*-values			Percent On-Target *t*-values		
	Int.	Piece 1	Piece 2	Int.	Piece 1	Piece 2
Working Alliance: Consultant Report	3.12**	2.05	.68	.12	.39	.09
Working Alliance: Clinician Report	1.15	1.19	-.79	.27	.97	1.08
Perceived Organizational Support	.77	2.30*	2.54*	-.23	-.76	-.85
Dose: Total # FF-Consultation Sessions	-.33	-1.75	-1.15	1.80	2.79*	2.53*
Dose: Ave. Duration of Consultation Sessions (Minutes)	-.46	1.02	1.09	-.67	-.78	-1.35
Dose: Duration (Months) of Consultation	-1.45	-1.52	-1.74	2.08*	2.88**	3.60**
Dose: Frequency of Consultation (Sessions/Month)	1.34	-.93	.16	.21	.98	1.05
# Consultation Sessions Prior to ABC Start	.97	.25	.70	-2.35*	-.07	.88

Note. Int. = Intercept. * *p* < .05; ** *p* < .01

Clinicians' Coding Reliability

Growth in Reliability Over Time

Table 8 shows the group level ICCs for behavior and comment coding

variables across the four 2.5-month periods of consultation. Table 8 also shows expert

coder ICCs from the current sample, and the correlations of clinicians' ICCs with time

(coded 1-4). Although only one correlation was significant because of the small

sample size (n = 4 time periods), 8 of the 10 correlations were above .5. In addition to

exploring growth in this way, individual clinicians' ICCs were modeled across time

using hierarchical linear modeling. As shown in Table 9, clinicians demonstrated

growth across the consultation period in coding following the lead, nurturance, and

delight. Additionally, behavior and comment coding ICCs were averaged into

summary measures, and on both of these measures, clinicians showed growth in

reliability over time.

Table 8   Clinicians' Coding Reliability: Group Level ICCs, Expert Comparison, and Correlation with Time

	Training Period					
Variable	1	2	3	4	Expert	*r*
Parent Behavior Coding						
Following the Lead	.80	.88	.89	.88	.80	.77
Not Following the Lead	.47	.62	.67	.49	.64	.15
Delight	.67	.69	.84	.76	.85	.70
Nurturing	.77	.76	.82	.80	.47	.70
Not Nurturing	.46	.67	.65	.80	.88	.92
Frightening	.64	.57	.18	.36	.91	-.76
Average Behavior Coding	*.50*	*.61*	*.62*	*.65*	--	*.89*
Comment Coding						
On-Target Comment Frequency	.91	.91	.95	.93	.84	.67
Pct. of On-Target Comments	.45	.55	.59	.67	.65	.99*
Percent of Missed Opportunities	.40	.39	.33	.53	.72	.51
Average Components	.56	.71	.76	.76	.74	.89
Average Comment Coding	*.47*	*.57*	*.57*	*.63*	--	*.92*

Note. Although nearly all ABC sessions used in consultation were coded by both the consultant and the clinician, the Single Measures ICC is used for more appropriate comparison with the Expert ICC. * $p < .05$

Table 9   Clinicians' Coding Reliability: Linear Growth Models of Change over Time

Variable	Coefficient	SE	t-ratio	p-value
Parent Behavior Coding				
Following the Lead	.04	.02	2.86	.008
Not Following the Lead	.03	.02	1.47	.16
Delight	.09	.03	2.78	.01
Nurturing	.07	.02	2.83	.009
Not Nurturing	.01	.04	.22	.83
Frightening	-.01	.05	-.20	.85
Average Behavior Coding	*.05*	*.02*	*3.21*	*.004*
Comment Coding				
On-Target Comment Frequency	.03	.02	1.61	.12
Percent of On-Target Comments	.08	.04	2.02	.054
Percent of Missed Opportunities	.05	.03	1.69	.10
Average Components	.04	.02	1.59	.12
Average Comment Coding	*.05*	*.02*	*2.96*	*.006*

Coding Reliability as a Predictor of Fidelity

As shown in Table 10, clinicians' behavior coding reliability in the first 2.5 months of consultation was associated with a higher comment frequency, $\beta_{01} = .59$, t (25) = 2.90, $p < .01$, and a higher percentage of on-target comments, $\beta_{01} = 1.25$, t (25) = 2.32, $p < .05$, when beginning to implement ABC (i.e., intercept). Clinicians' comment coding reliability in the first 2.5 months of consultation was not associated with fidelity intercepts, but was associated with stronger growth in percentage of on-target comments during the first 5 months of consultation, $\beta_{11} = .31$, t (25) = 2.65, $p < .05$. Clinicians' comment coding reliability during the first 5 months of consultation was associated with more rapid growth in percentage of on-target comments over both

the first, β_{11} = .32, t (25) = 2.20, p < .05, and second 5 months of consultation, β_{21} = .50, t (25) = 2.18, p < .05.

Table 10   Clinicians' Coding Reliability: Predicting Fidelity Growth and Intercepts

Coding Reliability Variables	Comment Rate t-values			% On-Target t-values		
	Int.	Piece 1	Piece 2	Int.	Piece 1	Piece 2
Comment Reliability – First 2.5 months	-.10	.63	1.34	-.49	-2.65*	-1.84
Comment Reliability – First 5 months	--	1.30	1.13	--	-2.20*	-2.18*
Behavior Reliability – First 2.5 months	2.90*	.77	.26	-2.32*	-1.75	-1.95
Behavior Reliability – First 5 months	--	.33	.25	--	-1.41	-.63

Note. Int. = Intercept. * p < .05

Predictors of Growth in Coding Reliability

As shown in Table 11, three variables predicted clinicians' growth in behavior coding reliability during the consultation period. First, consultants' setting of coding goals predicted slower growth in behavior coding reliability across the consultation period, β_{11} = -.50, t (25) = -3.53, p < .01. Clinicians reporting stronger working alliance with consultants also showed weaker growth in behavior coding reliability, β_{11} = -.12, t (25) = -2.68, p < .05. Conversely, clinicians who reported lower organizational support showed more rapid growth in behavior coding reliability, β_{11} = -.05, t (24) = -3.02, p < .01. In addition, clinicians' self-directed positive coding

74

feedback predicted stronger growth in comment coding reliability during consultation,

$\beta_{11} = .61$, $t\,(25) = 2.20$, $p < .05$.

Table 11   Clinicians' Coding Reliability: Predictors of Growth in Reliability over Time

Predictor Variable	Behavior Coding Reliability Growth t	Comment Coding Reliability Growth t
Consultant Variables		
Coding Feedback	1.17	-1.39
Coding Feedback – Single Behavior	.89	-1.70
Coding Feedback – 5 Minute Clip	1.63	-.07
Coding Feedback – Neutral	.96	-1.37
Coding Feedback – Positive	-1.70	1.04
Coding Goal Setting	3.53**	.84
Coding Goal Setting – Most Specific	1.84	-.01
Live Coding	-.70	-.40
Clinician Variables		
Coding Feedback	.85	.18
Coding Feedback – Single Behavior	.21	-.06
Coding Feedback – 5 Minute Clip	1.10	.63
Coding Feedback – Neutral	.94	.73
Coding Feedback – Positive	.61	-2.20*
Live Coding	-1.30	-1.49
Silence while Watching Video	-1.35	-1.05
Alternative Predictors		
Working Alliance: Consultant Report	1.39	.27
Working Alliance: Clinician Report	-2.68*	.49
Perceived Organizational Support	3.02**	.11
Dose: Total Number FF-Consultation Sessions	-.24	-1.43
Dose: Average Duration of Consultation Sessions	.74	.66
Dose: Duration (Months) of Consultation	.43	-1.37
Dose: Frequency of Consultation (Sessions/Month)	-.60	-.45
Number Consultation Sessions Prior to ABC Start	-.59	.34

Note. * $p < .05$; ** $p < .01$

Discussion

This study extended results from Study 1 by replicating findings of fidelity growth over the course of ABC consultation. As in Study 1, traditional group consultation and fidelity-focused consultation occurred simultaneously; unlike Study 1, however, the majority of clinicians (all but the six Study 1 participants) began both types of consultation at the same time. In Study 1, prior to initiation of fidelity-focused consultation, minimal change in fidelity was observed during the first few months of ABC implementation. In the current study, fidelity growth began immediately after clinicians began implementing ABC and participating in consultation. This finding lowers the likelihood that order/timing effects drove Study 1 findings, and makes it appear unlikely that receiving traditional consultation for several months was a necessary prerequisite for observing the rapid response to fidelity-focused consultation.

At the same time, Study 1 raised questions about the necessary dose of fidelity-focused consultation, because fidelity growth was observed only during the first 5 months of fidelity-focused consultation. The current study argues against reducing the current amount of ABC training and consultation, because clinicians continued to improve in two of the four fidelity outcomes during the second 5 months of consultation. Comparing intercepts of fidelity outcomes, Study 2 clinicians appeared to begin implementing ABC with stronger fidelity than Study 1 clinicians. This discrepancy may relate to true differences in the aptitude of the samples of clinicians, or may reflect improved ABC training. Specifically, between 2012 and 2015, initial

ABC training workshops increasingly focused on fidelity coding, and evolved to include active learning strategies such as responding to video-based coding questions with iClickers. Additionally, clinicians now begin fidelity-focused consultation before they begin implementing ABC, and receive additional training in coding and commenting in these individualized sessions. In the current study, clinicians participated in, on average, 2.5 fidelity-focused consultation sessions ($SD = 2.6$, range: $0 - 8$) prior to beginning to implement ABC, and the number of these prior sessions was associated with stronger initial percentage of on-target comments. Thus, differences in training methods make it difficult to interpret direct comparisons in absolute levels of fidelity between Study 1 and Study 2.

This study also demonstrated growth in clinicians' coding reliability during consultation, and adds to a currently very small literature on change in clinicians' ability to make self-ratings over time. The results are exciting because much of the literature has cast doubt on clinicians' ability to self-rate performance. In the current study, not only did clinicians show growth in coding reliability, but as a group, by the end of consultation, their interrater reliability with consultants was comparable to that of expert coders for many scales. These findings suggest that the sustainability shown in Study 1 may be in part due to clinicians becoming reliable self-assessors with a tool that they can continue to use for self-supervision.

A primary goal of the current study was to examine components of consultation as predictors of future fidelity in two ways: from one consultation session to the next ABC session, and across time. In support of feedback theories,

consultants' provision of neutral or negatively-valenced feedback that compared

commenting in a 5-minute clip with certification criteria was associated with the

clinician making more frequent comments in the following ABC session. This is

consistent with Feedback Intervention Theory, which asserts that providing feedback

that an individual has not met clear goals or standards will lead the individual to strive

to improve performance (Kluger & DeNisi, 1996, 1998). Also consistent with

Feedback Intervention Theory, which asserts that providing feedback that an

individual has exceeded goals or standards may lead to decreased effort and

performance, positively-valenced feedback comparing commenting with certification

criteria was not associated with future fidelity. Somewhat inconsistent with Feedback

Intervention Theory was the level of feedback specificity associated with higher

fidelity in the following session, the 5-minute clip, or the middle level of specificity

coded in the current study. According to Feedback Intervention Theory, highly

specific feedback is most likely to lead to behavior change because it directs attention

to lower-level task-learning processes (Kluger & DeNisi, 1996). In the current study,

the most specific level of feedback (i.e., single comments) was not associated with

future fidelity. One factor influencing this finding is that certification criteria is

defined at the level of the 5-minute clip, so nearly all feedback referencing

certification criteria was provided at this level. However, coding feedback at the level

of the 5-minute clip was also linked to more frequent comments in the next ABC

session. Although this association between coding feedback and future fidelity did not

hold in hierarchical analyses, it provides some additional support for the efficacy of a middle level of specificity in feedback.

In addition to comment feedback referencing certification criteria, comment feedback providing numbers without any other reference was also associated with higher comment frequency in the next ABC session. This finding appears to conflict with Feedback Intervention Theory, which claims that "outcome feedback interventions" or "mere knowledge of results" may impede learning of complex tasks (Kluger & DeNisi, 1996). However, Feedback Intervention Theory also suggests that feedback is often compared with multiple standards, and the resulting feedback-standard discrepancies are combined into an overall feedback evaluation. Thus, it is likely that when consultants provide clinicians with numbers-based feedback on fidelity, clinicians compare these numbers with various internal standards, such as their past performance, goals, or certification criteria. Numbers-based feedback may thus take on whatever reference or standard is most relevant to the clinician at that time, and may be more motivating because it is personalized to the clinician's internal standards.

Another finding consistent with Feedback Intervention Theory was that consultants' setting of commenting goals was associated with accelerated growth in clinicians' percentage of on-target comments during the first 5 months of consultation. Kluger and DeNisi (1996) assert that goal setting directs attention to lower-level task learning processes, making behavior change more likely. Consistent with this theory, only more specific goals, which were more likely to activate task-learning processes

(Kluger & DeNisi, 1996), were associated with growth in percentage of on-target comments; less specific goals were not.

Consultants' setting of coding goals also predicted fidelity in both session-to-session and linear growth models. Specifically, setting of coding goals was associated with lower comment frequency, and trended toward an association with lower percentage of on-target comments, in the next session. Consultants' setting of highly specific coding goals was also associated with slower growth in clinicians' comment frequency over the first 5 months of consultation. One interpretation is that setting coding goals may have directed clinicians' attention away from commenting fidelity, and interfered with improving fidelity to ABC. However, another interpretation is that clinicians who were relatively weak at coding had consultants set coding goals for them, and these clinicians also showed slower growth in fidelity. This interpretation is supported by the association of clinicians' early comment coding reliability with slower growth in fidelity (specifically, percentage of on-target comments) over both the first and second half of consultation. It is also supported by the association between consultants' setting of coding goals and slower growth in behavior coding reliability over the course of consultation, suggesting that setting of coding goals may have been a responsive technique used with clinicians who were struggling significantly with both coding and commenting. Relatedly, clinicians' self-provision of positively-valenced coding feedback was associated with greater growth in comment frequency over the first 5 months of consultation. This finding seems to represent the opposite side of the coin: clinicians who were successful at coding

tended to recognize it and give themselves positive feedback, and also showed greater growth in fidelity.

Mixed results were found for client outcome feedback, and no bivariate correlations held in HLM. On the one hand, consultants' linking of client progress to fidelity was associated with clinicians making fewer comments in the next ABC session. On the other hand, consultants' provision of comment feedback that positively described the impact of comments on the client was associated with a higher percentage of on-target comments in the next ABC session. These are similar consultant behaviors, and the focus of the consultant's statement (i.e., the client or the clinician) would determine whether it would be coded as discussion of client progress or comment feedback. It is possible that these different foci moderated the impact of client outcome feedback, with comment feedback making the link to the clinician's own behavior more explicit and more motivating to the clinician than feedback focused on client progress. It would be interesting to understand if consultants used different levels of specificity in client progress discussion than in comment feedback. However, most research about provision of client outcome feedback links it to improved outcomes for clients (Bickman et al., 2011; Worthen & Lambert, 2007), so perhaps it is not surprising that clear links to clinician fidelity were not found.

With regard to alternative predictors, several interesting findings emerged. Unlike prior work linking dose with greater fidelity (Beidas et al., 2012; Schwalbe et al., 2014), in the current study, dose was associated with slower growth in percentage of on-target comments. Overall, dose in the current study was much higher compared

with Beidas et al.'s (2012) study, in which the average participant received 7 hours of consultation, and with most studies of motivational interviewing consultation, which provided a median of 5 hours of consultation (Schwalbe et al., 2014). Further, in Beidas et al. (2012)'s study, consultation occurred in groups, whereas in the current study, it was individualized, and consultants who struggled more with learning ABC were allowed to receive additional consultation to reach certification criteria. These methodological differences may explain differences in findings.

Stronger consultant-reported working alliance was associated with higher comment frequency when clinicians began implementing ABC. This finding may reflect that clinicians who began ABC with some skill were easier for consultants to work and develop rapport with. However, consultant-reported alliance also trended toward predicting growth in commenting during the first 5 months of consultation. Because alliance was assessed only once, at the end of the consultation period, the direction of effects is unclear. Alliance may have supported growth in fidelity, or growth in fidelity may have helped build a strong working alliance. In support of the second possibility, in studies of client-therapist working alliance, some research suggests that therapist-rated working alliance, particularly when rated later in therapy, is subject to "reverse causation," and is influenced by therapeutic outcome (Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz & Gallop, 2011; Kivlighan & Shaughnessy, 1995). Although it is likely that both directions of effects contribute to findings, future studies of working alliance would need to measure alliance at multiple points during consultation to parse out these contributions.

Surprisingly, clinician-rated working alliance was associated with slower growth in behavior coding reliability over time. One possible explanation is that higher initial reliability was associated with stronger working alliance, and these clinicians showed slower growth in reliability because they started higher and had less room to improve; however, this hypothesis can be ruled out because clinician-rated alliance was not associated with the behavior coding reliability intercept. Another possible explanation is that when working alliance was strong, clinicians felt comfortable, and felt less need to make a strong effort at their weekly coding assignments. It is also possible that this finding is a Type I error; it is likely that some of the effects reported in the current study are Type I errors, due to the large number of analyses run. However, Schoenwald et al. (2004) found poorer fidelity and client outcomes when clinician-reported alliance was strong, so the finding is not unprecedented in the literature.

Another surprising set of findings involved perceived organizational support. Specifically, clinicians who reported lower agency support showed stronger growth in comment frequency across both the first and second halves of consultation, as well as stronger growth in behavior coding reliability across the consultation period, as compared with clinicians with higher agency support. Perhaps clinicians who felt less supported by their agencies were more receptive and responsive to the support provided through ABC consultation, and thus demonstrated greater growth in fidelity and coding reliability. One study of implementation support with fidelity monitoring found improved staff retention in the community agencies randomized to receive

implementation support (Aarons, Sommerfeld, Hecht, Silovsky, & Chaffin, 2009).

Together, these findings suggest that external support through consultation may

compensate for poor agency support, enhance implementation, and reduce burnout.

Overall, in examining predictors of change in fidelity and coding reliability, no

single component clearly stood out as a consistent predictor across different outcomes

and time periods. This may be a result of the complexity of the multicomponent

package of fidelity-focused consultation; of Milne et al.'s (2008) list of 26 effective

supervisory practices, over 10 were identified and coded for this study. Given that so

many techniques were used, it may be that if a consultant does not use one active

ingredient of consultation, they may compensate with other active ingredients,

obscuring effects. A number of strategies measured in the current study may be

effective.

Measurement issues may also have contributed to the inconsistent effects

observed in the current study. Specifically, several variables of interest, particularly

those coded for clinicians, had low reliability and were not examined as potential

predictors of clinicians' growth during consultation. In session-to-session models, the

majority of effects were observed for comment frequency, and not percentage of on-

target comments. In the current study, reliability for percentage of on-target

comments (ICC = .66) was somewhat lower than previous work (e.g., Study 1 ICC =

.76), as was reliability of comment frequency (Study 2 ICC = .84 vs. Study 1 ICC =

.95). Lower reliability may have made Type II errors more likely, particularly in the

session-to-session models, in which there was no data aggregation as there was in

linear growth models. Another possible measurement issue was that percentage of the full consultation session was selected as the way to examine the dose of different consultation strategies. Overall time (unadjusted for the length of the consultation session) or rate (i.e., number of times per consultation session or per minute) are other ways to examine the dose of consultation strategies. In the current study, these alternate forms of measurement were excluded to limit the number of analyses conducted; however, it would be interesting to compare results between measurement types. Future work with session-level analyses should also control for the clinician's prior fidelity, which is likely to be correlated with next-session fidelity. Finding associations between components of consultation and next-session fidelity, after controlling for prior fidelity, would provide stronger evidence for the directionality of consultation strategies leading to fidelity growth.

Alternatively, it is also possible that the consultation strategies studied were simply not very effective in helping clinicians to improve fidelity. Kluger and DeNisi (1996) found that verbal feedback interventions attenuated effects, and theorized that feedback that is verbally delivered highlights the person giving the feedback, which activates the recipient's own higher-level self-oriented processes, and thereby inhibits change. Perhaps the verbal behaviors coded from consultation sessions generally were not effective. In this case, the growth in fidelity observed in the current study may relate to unmeasured processes that occur outside of consultation sessions, such as the process of completing self-coding, receiving and reviewing the consultant's coding, and comparing this nonverbal feedback with one's own goals or standards.

That said, many of the results related to coding. One of the more interesting and novel findings of the current study was the association between coding reliability and concurrent and future ABC fidelity. Baseline coding ability when beginning ABC appeared to tap clinicians' understanding of the intervention because it predicted initial fidelity to the model. However, coding ability also predicted growth in fidelity over time. Because coding ability and fidelity both increased during the consultation period, improving coding may be a mechanism by which clinicians improve fidelity. These results support the integration of self-coding into consultation procedures, a component of consultation that is rare, with a few exceptions (e.g., Garcia et al., 2017; Isenhart et al., 2014). Although the current evidence for self-coding as an active ingredient of consultation is correlational, future work could test the impact of self-coding in a group design, in which clinicians are assigned to two types of fidelity-focused consultation, one in which they code themselves and receive feedback from an expert coder, and one in which they just receive coding and feedback from an expert coder.

To summarize Study 1 and Study 2 within the context of an implementation framework, Fixsen et al. (2005) proposed that core intervention practices are taught to clinicians through communication link processes. ABC's core intervention practice, in the moment commenting, was identified prior to the current studies. In Study 1, fidelity-focused consultation was established as a communication link process and core component of implementation through its observed impact on clinicians' in the moment commenting, over and above traditional consultation. In Study 2, learning to

code one's own fidelity was identified as a critical subcomponent of fidelity-focused consultation. One of the unique aspects of Fixsen et al.'s (2005) model of implementation is the specification of a feedback loop in which fidelity is regularly measured to provide feedback to consultants and trainers. The current study suggests that measurement of fidelity, and indeed, self-assessment of fidelity, can provide a feedback loop to clinicians themselves. Fixsen et al. (2005) proposed that these implementation processes occur within a sphere of influence shaped by an organization's structure and culture. The Study 2 findings on organizational support emphasize the impact of this sphere of influence. Finally, Fixsen et al. (2005) defined stages of implementation, including exploration, initial implementation, full operation, and long-term sustainability. One of the strengths of the current studies is that the impact of fidelity-focused consultation was demonstrated across multiple stages, from initial implementation through sustainability.

Together, the two studies join a fairly small literature showing that consultation procedures that incorporate fidelity coding feedback are effective in promoting implementation outcomes (Martino et al., 2016; Miller et al., 2004; Weck et al., 2017). They join an even smaller literature that suggests added benefit for self-observation and self-coding of fidelity (Garcia et al., 2017; Isenhart et al., 2014). To my knowledge, the current work is unique in demonstrating a link between clinicians' self-coding reliability and their growth in fidelity, and thereby suggests a potential mechanism by which consultation can impact implementation.

REFERENCES

Aarons, G. A., Sommerfeld, D. H., Hecht, D. B., Silovsky, J. F., & Chaffin, M. J. (2009). The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: evidence for a protective effect. *Journal of Consulting and Clinical Psychology*, *77*, 270-280. http://dx.doi.org/10.1037/a0013223

Andrzejewski, M. E., Kirby, K. C., Morral, A. R., & Iguchi, M. Y. (2001). Technology transfer through performance management: The effects of graphical feedback and positive reinforcement on drug treatment counselors' behavior. *Drug and Alcohol Dependence*, *63*, 179-186. https://doi.org/10.1016/S0376-8716(00)00207-6

Bahrick, A. S. (1989). *Role induction for counselor trainees: Effects on the supervisory working alliance* (Doctoral dissertation). Retrieved from OhioLINK Electronic Theses & Dissertations Center.

Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., ... & 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*, 518-529. 10.1007/s10488-013-0485-5

Beidas, R. S., Edmunds, J. M., Cannuscio, C. C., Gallagher, M., Downey, M. M., & Kendall, P. C. (2013). Therapists' perspectives on the effective elements of consultation following training. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*, 507-517. 10.1007/s10488-013-0475-7

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

Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., ... & Adams, D. R. (2015). Individual and organizational factors related to community clinicians' use of therapy techniques in a large public mental health system. *JAMA Pediatrics*, *169*, 374-382. 10.1001/jamapediatrics.2014.3736

Bernard, K., Dozier, M., Bick, J., & Gordon, M. K. (2015). Intervening to enhance cortisol regulation among children at risk for neglect: results of a randomized clinical trial. *Development and Psychopathology*, *27*, 829-841. http://dx.doi.org/10.1017/S095457941400073X

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*, 623-636. http://dx.doi.org/10.1111/j.1467-8624.2011.01712.x

Bernard, K., Hostinar, C. E., & Dozier, M. (2015). Intervention effects on diurnal cortisol rhythms of child protective services–referred infants in early childhood: Preschool follow-up results of a randomized clinical trial. *JAMA Pediatrics*, *169*, 112-119. 10.1001/jamapediatrics.2014.2369

Bernard, K., Simons, R., & Dozier, M. (2015). Effects of an attachment‑based intervention on child protective services–referred mothers' event‑related potentials to children's emotions. *Child Development*, *86*, 1673-1684.

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services*, *62*, 1423-1429. https://doi.org/10.1176/appi.ps.002052011

Caron, E. B., Bernard, K., & Dozier, M. (2016). In vivo feedback predicts parent behavior change in the Attachment and Biobehavioral Catch-up intervention. *Journal of Clinical Child & Adolescent Psychology*. Advance online publication. http://dx.doi.org/10.1080/15374416.2016.1141359

Caron, E. B., Weston-Lee, P., Haggerty, D., & Dozier, M. (2016). Community implementation outcomes of Attachment and Biobehavioral Catch-up. *Child Abuse & Neglect*, *53*, 128-137. https://doi.org/10.1016/j.chiabu.2015.11.010

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*, 307-320.

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., ... & Starace, N. (2011). Evidence‑based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, *18*, 154-172. http://dx.doi.org/10.1111/j.1468-2850.2011.01247.x

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. American Journal of Mental Deficiency, 127–137.

Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology*, *79*, 267-278. http://dx.doi.org/10.1037/a0023668

Dennin, M. K., & Ellis, M. V. (2003). Effects of a method of self-supervision for counselor trainees. *Journal of Counseling Psychology*, *50*, 69-83. http://dx.doi.org/10.1037/0022-0167.50.1.69

Dunn, C., Darnell, D., Atkins, D. C., Hallgren, K. A., Imel, Z. E., Bumgardner, K., ... & 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. https://doi.org/10.1016/j.jsat.2016.02.008

Edmunds, J. M., Beidas, R. S., & Kendall, P. C. (2013a). Dissemination and implementation of evidence–based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice*, *20*, 152-165. http://dx.doi.org/10.1111/cpsp.12031

Edmunds, J. M., Kendall, P. C., Ringle, V. A., Read, K. L., Brodman, D. M., Pimentel, S. S., & Beidas, R. S. (2013b). 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*, 456-466. http://dx.doi.org/10.1007/s10488-013-0490-8

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature (FMHI Publication No. 231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Friedlander, M. L., Ellis, M. V., Siegel, S. M., Raymond, L., Haase, R. F., & Highlen, P. S. (1988). Generalizing from segments to sessions: Should it be done? Journal of Counseling Psychology, 35, 243–250. 10.1037/0022-0167.35.3.243

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*, 730-741. http://dx.doi.org/10.1080/15374416.2014.910790

Garcia, I., James, R. W., Bischof, P., & Baroffio, A. (2017). Self-observation and peer feedback as a faculty development approach for problem-based learning tutors: A program evaluation. *Teaching and Learning in Medicine*. Advance online publication.

Garland, A., & Schoenwald, S. K. (2013). Use of effective and efficient quality control methods to implement psychosocial interventions. *Clinical Psychology: Science and Practice*, *20*, 33-43. http://dx.doi.org/10.1111/cpsp.12021

Goodyear, R. K. (2014). Supervision as pedagogy: Attending to its essential instructional and learning processes. *The Clinical Supervisor*, *33*, 82-99. http://dx.doi.org/10.1080/07325223.2014.918914

Henggeler, S. W., Chapman, J. E., Rowland, M. D., Halliday-Boykins, C. A., Randall, J., Shackelford, J., & Schoenwald, S. K. (2008a). Statewide adoption and initial implementation of contingency management for substance-abusing adolescents. *Journal of Consulting and Clinical Psychology*, *76*(4), 556-567. 10.1037/0022-006X.76.4.556

Henggeler, S. W., Sheidow, A. J., Cunningham, P. B., Donohue, B. C., & Ford, J. D. (2008b). Promoting the implementation of an evidence-based intervention for adolescent marijuana abuse in community settings: Testing the use of intensive quality assurance. *Journal of Clinical Child & Adolescent Psychology*, *37*, 682-689. http://dx.doi.org/10.1080/15374410802148087

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review*, *30*, 448-466. http://dx.doi.org/10.1016/j.cpr.2010.02.005

Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (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*, 229-243. 10.1007/s10488-014-0548-2

Holth, P., Torsheim, T., Sheidow, A. J., Ogden, T., & Henggeler, S. W. (2011). Intensive quality assurance of therapist adherence to behavioral interventions for adolescent substance use problems. *Journal of Child & Adolescent Substance Abuse*, *20*, 289-313. http://dx.doi.org/10.1080/1067828X.2011.581974

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88-110. http://dx.doi.org/10.1080/19345740802539325

Hurlburt, M. S., Garland, A. F., Nguyen, K., & Brookman-Frazee, L. (2010). Child and family therapy process: Concordance of therapist and observational perspectives. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*, 230-244. 10.1007/s10488-009-0251-x

Imel, Z. E., Baldwin, S. A., Baer, J. S., Hartzler, B., Dunn, C., Rosengren, D. B., & Atkins, D. C. (2014). Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients. *Journal of Consulting and Clinical Psychology*, *82*, 472-481. http://dx.doi.org/10.1037/a0036158

Isenhart, 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*, 164-170. http://dx.doi.org/10.3109/14659891.2013.765514

Kivlighan, D. M., & Shaughnessy, P. (1995). Analysis of the development of the working alliance using hierarchical linear modeling. *Journal of Counseling Psychology*, *42*, 338-349. http://dx.doi.org/10.1037/0022-0167.42.3.338

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284. http://dx.doi.org/10.1037/0033-2909.119.2.254

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, *7*, 67-72. http://dx.doi.org/10.1111/1467-8721.ep10772989

Lewis-Morrarty, E., Dozier, M., Bernard, K., Terracciano, S. M., & Moore, S. V. (2012). Cognitive flexibility and theory of mind outcomes among foster children: Preschool follow-up results of a randomized clinical trial. *Journal of Adolescent Health*, *51*, S17-S22. http://dx.doi.org/10.1016/j.jadohealth.2012.05.005

Lind, T., Bernard, K., Ross, E., & Dozier, M. (2014). Intervention effects on negative affect of CPS-referred children: Results of a randomized clinical trial. *Child Abuse & Neglect*, *38*, 1459-1467. http://dx.doi.org/10.1016/j.chiabu.2014.04.004

Lind, T., Raby, K. L., Caron, E. B., Roben, C. K., & Dozier, M. (2017). Enhancing
executive functioning among toddlers in foster care with an attachment-based
intervention. *Development and Psychopathology*, *29*, 575-586.
https://doi.org/10.1017/S0954579417000190

Loades, M. E., & Myles, P. J. (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*. https://doi.org/10.1017/S1754470X16000027

Lyon, A. R., Pullmann, M. D., Whitaker, K., Ludwig, K., Wasse, J. K., & McCauley,
E. (2017). A digital feedback system to support implementation of
measurement-based care by school-based mental health clinicians. *Journal of
Clinical Child & Adolescent Psychology*. Advance online publication.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance
with outcome and other variables: A meta-analytic review. *Journal of
Consulting and Clinical Psychology*, *68*, 438-450.
http://dx.doi.org/10.1037/0022-006X.68.3.438

Martino, S., Ball, S., Nich, C., Frankforter, T. L., & Carroll, K. M. (2009).
Correspondence of motivational enhancement treatment integrity ratings
among therapists, supervisors, and observers. *Psychotherapy Research*, *19*,
181-193. http://dx.doi.org/10.1080/10503300802688460

Martino, S., Paris, M., Añez, L., Nich, C., Canning-Ball, M., Hunkele, K., ... &
Carroll, K. M. (2016). The effectiveness and cost of clinical supervision for
motivational interviewing: a randomized controlled trial. *Journal of Substance
Abuse Treatment*, *68*, 11-23. https://doi.org/10.1016/j.jsat.2016.04.005

Meade, E. B., Dozier, M., & Bernard, K. (2014). Using video feedback as a tool in
training parent coaches: promising results from a single-subject design.
Attachment & Human Development, *16*, 356-370.
http://dx.doi.org/10.1080/14616734.2014.912488

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A
randomized trial of methods to help clinicians learn motivational interviewing.
Journal of Consulting and Clinical Psychology, *72*, 1050-1062.
http://dx.doi.org/10.1037/0022-006X.72.6.1050

Milne, D., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical
supervision work? Using a "best evidence synthesis" approach to construct a
basic model of supervision. *The Clinical Supervisor*, *27*, 170-190.
http://dx.doi.org/10.1080/07325220802487915

Milne, D., James, I., Keegan, D., & Dudley, M. (2002). Teacher's PETS: a new observational measure of experiential training interactions. *Clinical Psychology & Psychotherapy*, *9*, 187-199. 10.1002/cpp.308

Milne, D. L., Pilkington, J., Gracie, J., & James, I. (2003). Transferring skills from supervision to therapy: A qualitative and quantitative N= 1 analysis. *Behavioural and Cognitive Psychotherapy*, *31*, 193-202. https://doi.org/10.1017/S1352465803002078

Moyers, T. B., Manuel, J. K., Wilson, P. G., Hendrickson, S. M., Talcott, W., & Durand, P. (2008). A randomized trial investigating training in motivational interviewing for behavioral health providers. *Behavioural and Cognitive Psychotherapy*, *36*, 149-162. https://doi.org/10.1017/S1352465807004055

Murray, H. (2017). Evaluation of a trauma-focused CBT training programme for IAPT services. *Behavioural and Cognitive Psychotherapy*. Advance online publication. https://doi.org/10.1017/S1352465816000606

Nadeem, E., Gleacher, A., & Beidas, R. S. (2013). Consultation as an implementation strategy for evidence-based practices across multiple contexts: Unpacking the black box. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*, 439-450. http://dx.doi.org/10.1007/s10488-013-0502-8

National Advisory Mental Health Council. (2001). *Blueprint for Change: Research on Child and Adolescent Mental Health. A Report by the National Advisory Mental Health Council's Workgroup on Child and Adolescent Mental Health Intervention Development and Deployment*. Bethesda, Maryland: National Institutes of Health, National Institute of Mental Health.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., ... & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, *34*, 87-106.

Peavy, K. M., Guydish, J., Manuel, J. K., Campbell, B. K., Lisha, N., Le, T., ... & 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*, 222-228. https://doi.org/10.1016/j.jsat.2014.05.008

Persson, J. E., Bohman, B., Forsberg, L., Beckman, M., Tynelius, P., Rasmussen, F., & Ghaderi, A. (2016). Proficiency in motivational interviewing among nurses in child health services following workshop and supervision with systematic feedback. *PloS one*, *11*, e0163624. https://doi.org/10.1371/journal.pone.0163624

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*, 431-451. http://dx.doi.org/10.1016/j.ecresq.2008.02.001

President's New Freedom Commission on Mental Health. (2004). *Report of the President's New Freedom Commission on Mental Health*. Retrieved from http://www.nami.org/Template.cfm?Section=Policy&Template=/ContentManagement/ContentDisplay.cfm&ContentID=16699.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Rhoades, L., Eisenberger, R., & Armeli, S. (2001). Affective commitment to the organization: The contribution of perceived organizational support. *Journal of Applied Psychology*, *86*, 825-836. http://dx.doi.org/10.1037/0021-9010.86.5.825

Richman, G. S., Riordan, M. R., Reiss, M. L., Pyles, D. A., & Bailey, J. S. (1988). The effects of self-monitoring and supervisor feedback on staff performance in a residential setting. *Journal of Applied Behavior Analysis*, *21*, 401-409. http://dx.doi.org/10.1901/jaba.1988.21-401

Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child and Adolescent Psychiatric Clinics of North America*, *14*, 241-254. http://dx.doi.org/10.1016/j.chc.2004.05.002

Roben, C.K.P., Dozier, M., Caron, E., & Bernard, K. (in press). Moving an evidence-based parenting program into the community. *Child Development*.

Santa Ana, E. J., Martino, S., Ball, S. A., Nich, C., Frankforter, T. L., & Carroll, K. M. (2008). What is usual about "treatment-as-usual"? Data from two multisite effectiveness trials. *Journal of substance abuse treatment*, *35*, 369-379. http://dx.doi.org/10.1016/j.jsat.2008.01.003

Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, *77*, 410-421. http://dx.doi.org/10.1037/a0013788

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (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*, 94-104. http://dx.doi.org/10.1207/S15374424JCCP3301_10

Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research*, *5*, 223-239. http://dx.doi.org/10.1023/A:1026229102151

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In B.J. Zimmerman & D.H. Schunk (Eds.), *Self-regulated Learning and Academic Achievement* (pp. 83-110). New York, NY: Springer.

Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta‑analysis of training studies. *Addiction*, *109*, 1287-1294. 10.1111/add.12558

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, *73*, 106-115. http://dx.doi.org/10.1037/0022-006X.73.1.106

Tabak, R. G., Khoong, E. C., Chambers, D. A., & Brownson, R. C. (2012). Bridging research and practice: models for dissemination and implementation research. *American Journal of Preventive Medicine*, *43*, 337-350. https://doi.org/10.1016/j.amepre.2012.05.024

Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*, *78*, 144-155. http://dx.doi.org/10.1037/a0018544

Weck, F., Kaufmann, Y. M., & Höfling, V. (2017). Competence feedback improves CBT competence in trainee therapists: A randomized controlled pilot study. *Psychotherapy Research*, *27*, 501-509. http://dx.doi.org/10.1080/10503307.2015.1132857

Wehby, J. H., Maggin, D. M., Moore Partin, T. C., & 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, J. R., Ugueto, A. M., Herren, J., Afienko, S. R., & Rutt, C. (2011). Kernels vs. ears, and other questions for a science of treatment dissemination. *Clinical Psychology: Science and Practice*, *18*, 41-46. 10.1111/j.1468-2850.2010.01233.x

Worthen, V. E., & Lambert, M. J. (2007). Outcome oriented supervision: Advantages of adding systematic client tracking to supportive consultations. *Counselling and Psychotherapy Research*, *7*, 48-53. 10.1080/14733140601140873

Appendix A

IRB APPROVAL LETTERS

DATE: February 16, 2016

TO: Elizabeth Meade, BA
FROM: University of Delaware IRB

STUDY TITLE: [837682-1] Assessing Strategies for Fidelity-Focused Supervision

SUBMISSION TYPE: New Project

ACTION: APPROVED
APPROVAL DATE: February 16, 2016
EXPIRATION DATE: February 15, 2017
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # *(6,7)*

Thank you for your submission of New Project materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that <u>informed consent</u> is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Nicole Farnese-McFarlane at (302) 831-1119 or nicolefm@udel.edu. Please include your study title and reference number in all correspondence with this office.

DATE: February 3, 2017

TO: Elizabeth Caron, MA
FROM: University of Delaware IRB

STUDY TITLE: [837682-2] Assessing Strategies for Fidelity-Focused Supervision

SUBMISSION TYPE: Continuing Review/Progress Report

ACTION: APPROVED
APPROVAL DATE: February 3, 2017
EXPIRATION DATE: February 15, 2018
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # *(6,7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that <u>informed consent</u> is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Nicole Farnese-McFarlane at (302) 831-1119 or nicolefm@udel.edu. Please include your study title and reference number in all correspondence with this office.

FIDELITY-FOCUSED CONSULTATION CODING MANUAL

The first option that you will see when coding is the "participant" code.

Other		o
Supervisor		s
Parent Coach		p

Most of what you code will be under Supervisor or Parent Coach.

However, a few behaviors are coded under the **Other** participant code**.**

<u>Other Behaviors</u>

⊿ **Active Learning**			
AL Modeling		m	M
AL Role-play		r	R
AL Thinking/silence *cod...		t	T
AL Live coding		l	L
⊿ Other (not participant ...			
talk/silence related to te...		h	H
planning/scheduling/ag...		p	P
irrelevant ABC-/case rela...		i	I

<u>**Active Learning**</u>

<u>Only one</u> Active Learning behavior should be coded under the "Other" participant code (the others appear by default and there isn't a way to hide them):
Thinking/silence
  • **Thinking/silence** is what you will code when video is being watched, silently, during supervision; use the "*video continuous*" modifier for Realism)

Realism		
Video continuous		w
Video paused		x
PC's coding sheet		y
Worksheet		z
Other (e.g., spontaneous)		01

- You will use this instead of "Live Coding" when supervisors watch, pause, and then "live code."
- You will also use this when live coding with video turns into just silent watching (> 5-10 seconds of silence)
- You may also code "Thinking/Silence" if there are long pauses (>5 sec) as the coach looks at a worksheet or coding sheet and tries to come up with a comment. These would be coded with "*PC's coding sheet*" or "*Worksheet*" as the modifier.

Other (not participant specific)

- **Talk/silence related to tech issues**
    - Includes conversations beginning with "Can you hear me?", discussion of how weird/annoying the web conferencing software is, narration of screen sharing attempts, and periods of silence during disruptions in calls, etc.
    - Don't include time spent looking for files as a tech issue, unless it is explicitly discussed as such (e.g., "My email is so weird," etc.)
    - Don't code discussion of ABC session tech issues with this code (probably, code as Reflection/General ABC)
        - "I have this really old laptop, and it wouldn't play the session 3 videos, so I just tried to describe them instead."
        - "The video I sent you is only 20 minutes long, because my camera shut off. I'm still trying to figure out the battery life, it's kind of deceiving."

- **Planning/scheduling/agenda setting**
    - Includes discussion of how/when/which videos will be sent, when coding will be assigned, and scheduling of upcoming meetings; often includes discussion about when sessions with families are scheduled in order to plan upcoming coding and supervision sessions
    - Also including agenda setting statements, like "Let's look at the Case 1 Session 10 coding first."

- **Irrelevant ABC-/case-related talk**
    - Includes discussion of case details/stories that don't really contribute to a case conceptualization (e.g., discussion of random life events, like

child illness, leading to difficulties with scheduling; discussion of cute child behaviors)

- o Includes back-and-forth discussion of what case the coach is on with different families, and "which case is which" talk, without much detail that would fit in better as case discussion. The point of this talk often seems to be related to planning which sessions will be coded.
    - ▪ "So, the case with the grandmother who sits on the couch, where are you with them?" "I'm going to do session 7 on Monday." "And what about the case with twins?" "I'm supposed to do session 4 on Tuesday, but mom isn't sure she wants to continue ABC because she just got a new job." "Oh, okay. And you said you have 2 new cases?" "Yes, I did a pre-session with one family, and I'll be doing session 2 tomorrow with the other."
- o Discussion of general ABC dissemination progress (e.g., where else people are doing ABC)
- o Discussion of ABC-related response in community, how coach has been involved in systems-intervention (e.g., presenting to DHS/CPS, etc. on ABC)

Supervisor Behaviors

When you hit "s" to begin coding a supervisor behavior, the following options will come up:

Provision of feedback			
Comment Feedback		f	F
Coding Feedback		e	E
Questions/Prompts			
Q/P topic - comment		q	Q
Q/P topic - coding		a	A
Q/P topic - client		x	X
Q/P topic - other		o	O
Informing/Recommen...			
I/R topic - other		k	K
I/R topic - commenting		g	G
I/R topic - coding		c	C
Client Discussion			
Client discussion		d	D
Active Learning			
AL Modeling		m	M
AL Role-play		r	R
AL Thinking/silence *cod...		t	T
AL Live coding		l	L
Goal Setting			
Goal topic - comment		j	J
Goal topic - coding		u	U
Supporting/Understan...			
Supporting or Understa...		s	S
Other (participant-spe...			
Discussion of PC's workl...		w	W

Provision of feedback

There are two options for feedback, Comment Feedback and Coding Feedback.

Provision of feedback			
Comment Feedback		f	F
Coding Feedback		e	E

Note: when deciding between coding as "feedback" versus "informing/recommending," listen for a "you" referencing the parent coach (or "we" referencing both the supervisor and parent coach, more common in coding feedback). If you don't hear it, it's probably informing/recommending.

- One exception to the "you" rule is if the supervisor is sharing screens and is reading off the parent coach's coding sheet, or uses language like "too" or "as well" implying a comparison to the coach's coding. e.g., "19:27, child whimpers and caregiver says what's wrong," I had that as well.

Comment Feedback

Comment feedback is coded when the supervisor provides feedback about the coach's past commenting based on coding sheets, and when the supervisor provides feedback about practice comments the coach makes in session (e.g., during worksheet or live commenting activity).

When you hit "f" to code Comment Feedback, the following options will come up:

Comment Feedback	
Topic - frequency	L
Topic - level/components	M
Topic - on/off-target	N
Topic - negative comments	03
Topic - other	O
Valence - positive	P
Valence - negative	Q
Valence - neutral	R
Reference - norms/other PCs	S
Reference - certification criteria	T
Reference - PC's past performance	U
Reference - goals	V
Reference - ITM supervisor	W
Reference - developmental	X
Reference - just #s	10
Reference - client	04
Reference - other/unclear	Y
Specificity - single comment	Z
Specificty - in 5-min clip	00
Specificity - more general	7

One, and only one, option from each area (Topic, Valence, Reference, Specificity) should be coded.

Topics:
- **Frequency**: feedback on number of comments, or number of missed opportunities
- **Level/components**: feedback on aspects of comments related to level, including feedback on use (or lack of use) of specific components such as the "reciprocal behavior description" or outcomes
- **On-/off-target**: feedback on comments being on- or off-target
- **Negative comments**: feedback on making negative, scaffolding, or shaping comments (including feedback about appropriately not making such comments, in early sessions), "you could have" comments, "not not following" comments, and negative-to-positive comment balance
- **Other**: topics include commenting on specific behavior targets, other aspects of commenting (e.g., balancing comments between 2 caregivers), and when the topic is unclear (e.g., "You did awesome with commenting this week!")
  - ○ Includes generic "a comment" stuff ("You made a comment after that behavior")

Valence:
- **Positive:** when supervisor uses words such as good job, great, awesome
- **Negative:** when supervisor discusses coach performance, finds coach lacking, and uses agentic (i.e., coach is agent, doing the actions, with an intentional feel) language; coded relatively infrequently
  - ○ "You missed a lot of behaviors this week"
  - ○ "You made two off-target comments"
- **Neutral:** when supervisor provides numeric feedback without any comparison words, balances negative and positive feedback, uses less agentic language (e.g., "forgot," "accidentally") or provides excuses for coach:
  - ○ "Your comments were 100% on-target this week"
  - ○ "You made 3 comments in this clip"
  - ○ "So that comment was off-target, but it's actually great that you made it, because now we have the opportunity to talk about it."
  - ○ "So it was a really good comment, it was just that you made it at the wrong time."
  - ○ "There were a few differences on our coding sheets."
  - ○ "You didn't make any comments, but this was a session 1 and there were so many negative behaviors."
  - ○ "It looks like you forgot to put in 0's when you didn't make a comment"
  - ○ "Your comment at 19:07 was off-target because you accidentally included a nurturance outcome in a following the lead comment."

Reference:

- **Norms/Other PCs:** when supervisor compares coach's performance to the typical coach, or other coaches she has supervised
  - "You made 3 comments in this clip, which is an awesome start. A lot of coaches don't even make 1 comment in their first session."
  - "You missed mostly negative behaviors in the coding, which is pretty typical for coaches when they are starting out"

- **Certification Criteria:** when supervisor compares coach's performance to "what we want to see" (i.e., 1 comment/minute or 5 comments in 5 minutes, <50% missed opportunities, average level of 1 or higher, >80% on-target); include times when supervisors compare performance to these numbers without specifically identifying them as certification criteria
  - *Note*: in Hawaii videos, you will hear 3 being used for the level criterion (recall that Hawaii coding sheet uses 1-5 level system)
  - Include times when coach appears to have criteria incorrect but is stating numbers authoritatively as though they are the correct numbers
  - "You made 4 comments, which is so close to the 5 comments we are looking for"
  - "So you made 1 off-target comment, but you were overall 85% on-target, which is above the cut-off that we look for"

- **PC's past performance**: compares this week's performance to prior performance
  - "Your level this week was a 1.00, which is the same as your level from last week."
  - "You made one more comment than you did last week, awesome job."
  - "You noticed a lot more negative behaviors this week, good job."

- **Goals:** compares performance to goals set in prior supervision session
  - "You included an outcome in one of your comments, which was one of your goals."
  - "You made comments toward both mom and dad, which was something we talked about last week"

- **ITM Supervisor**: this is coded only for Coding Feedback; used when the coach's coding is compared to the supervisor's coding. This is the most frequent Reference code for Coding Feedback.
  - "I had a few behaviors that you didn't code."
  - "We coded the comment a little differently."

- **Developmental:** compares performance to expectations based on a developmental trajectory of parent coaching
  - "You did an awesome job for your first time coding!"
  - "You had three comments, which is great for only your second session!"

- *Note*: "Norms/Other PCs" may overlap with this category, and should <u>trump</u> this category. For example, "There were a lot of differences in our coding, which is totally typical when you're just starting to code" should be coded as "Ref – norms/other PCs"

- **Just #'s**: this is coded <u>only for Comment Feedback</u>; provides numeric values representing coach performance, but does not compare these numbers to certification criteria or any other reference group
    - If a feedback statement seems like it could be coded in this category, or another category, code it as the other category; the only category "Just #'s" trumps is "Other/Unclear"
    - "You had an average level of 1.35 in this clip."
    - "You made two negative comments in this clip!"
    - Good job, that was a level 2 comment."
    - "There was one off-target comment in this video."
    - "Every single comment was on-target." (similar to saying 100%)

- **Client**: this is coded <u>only for Comment Feedback</u>; discusses impact of comments on parent/child
    - "You made a great nurturance comment in this video, which is so important for this parent."
    - "You had two outcomes in your comments, which really reminds the mom of why the targets are important."
    - "Did you see what happened after you made the scaffolding comment? Mom's next 3 behaviors were following the lead!" (*Note*: if there were a pause between the question and the next sentence, as if the supervisor actually meant for the coach to respond to the question, the question would be coded under "Question/Prompt," discussed below)
    - "There were two comments where you used the word "letting," which is a really tricky rule we have. Basically, we don't want parents to think that following the lead is letting their child do whatever they want."

- **Other/Unclear**: includes some other reference, not discussed above; most often coded for vague/unclear feedback
    - "You did an awesome job with your coding."
    - "You made so many comments in this clip!"
    - (after coach makes comment in comment practicing activity) "Good, good job."

Specificity:
- **Single comment**: references one (to two) comment(s)
    - "You even had a level 3 comment in this clip!"
    - "Your comment at 19:07 was off-target because you accidentally included a nurturance outcome in a following the lead comment."
    - "There were two comments where you used the word "letting," which is a really tricky rule we have. Basically, we don't want parents to think that following the lead is letting their child do whatever they want."

- **In 5-Min Clip**: discusses performance in specific clips (can include multiple clips, as long as they are referenced specifically)
  - "In this clip, you had 7 comments, 100% on-target, and an average level of 1.2."
  - **"**You had over 1 comment per minute in both of the clips we coded this week!"
  - "Our coding matched up really well for this clip."
  - "The main difference in our coding for this clip was that I coded more delight than you did. Delight can be hard to see the first time you watch, so I recommend watching a second time to make sure you catch all the delight."

- **More General**: discusses performance more generally or in non-specific terms
  - "Your coding has gotten so much better!"
  - "You did a great job with commenting this week."

Coding Feedback

Coding Feedback	
Topic - behavior coding	11
Topic - comment coding	12
Valence - positive b	13
Valence - negative b	14
Valence - neutral b	15
Ref - norms/other PCs	16
Ref - certification criteria	17
Ref - PC past performance	18
Ref - goals	19
Ref - ITM supervisor	20
Ref - developmental	21
Ref - other/general	22
Specificity - single behavior	23
Specificity - in 5 min b	8
Specificity - more general b	9

Overall, the Valence, Reference and Specificity specifiers for Coding Feedback are very similar to those for Commenting Feedback, and for that reason, are described above.

The main difference in Coding Feedback specifiers relates to Topic.

Topic:
- **Behavior Coding:** feedback relates to coding parent behaviors
  - "You coded a few 2's that I didn't code, and these were all times when the parent was setting what I thought was an appropriate limit."
  - "I had a lot more behaviors, toward the end of the clip, like from 17:00 on, a lot of delight when the parent was giggling at how the kid kept going for the camera."
    - Note that the "you" for the PC is implicit here – but "more" implies a comparison with the PC.
- **Comment Coding**: feedback relates to coding parent coach comments (or lack of comments, e.g., putting in 0's for missed opportunities)
  - "You gave it a level 2, and I gave it a level 1. I think you were seeing both the behavior description and the target, but for the behavior description to be counted as a full component, it has to include both the child behavior and the parent behavior."
- If feedback is too general or unclear to code one or the other, <u>do not code</u> a Topic.
  - "We only had a couple differences in our coding this week."
  - "There's really not much to talk about with coding, because you did awesome."

Specificity – Single Behavior can again refer to 1 <u>to 2</u> behaviors; really, it gets at discussing specific behaviors
- "There were a couple times when the parent tickled the child, which I coded as a 6 and you coded as a 2. The reason I coded a 6 was that the child was distressed, and the parent seemed to be using tickling to cheer the child up, or distract them for their distress."

<u>Statements NOT clear enough to code as feedback</u>

If a supervisor says something that sounds like feedback, but isn't specific enough to categorize as Comment Feedback or Coding Feedback ("But really, you're doing great" or "Awesome job this week"), code it as "Supporting/Understanding" (discussed below).

If supervisor says something that is probably intended as feedback, but does not explicitly discuss coach performance, it should be instead coded as Informing/Recommending. This typically occurs when supervisors seem to be trying to avoid giving negative feedback; for example:
- "So using the word "letting" makes a following the lead comment off-target. Because we want parents to know that following the lead is more than just letting your child do what they want."
- "So another thing you'll see on my coding sheet is two 2's at the end. Whenever a parent doesn't pay attention to their child for a full minute, we code a 2."

Questions/Prompts

◢ Questions/Prompts			
Q/P topic - comment		q	Q
Q/P topic - coding		a	A
Q/P topic - client		x	X
Q/P topic - other		o	O

There are 4 types of questions/prompts that can be coded, based on topic:
- **Comment:** question relates to understanding of comment topic, or coach's experience with commenting
  - (explaining why comment was off-target) "Does that make sense?"
  - "How did you feel the pudding activity went, in terms of commenting?"
- **Coding:** question relates to understanding of coding topic (including coding of both behaviors or comments), coach's experience with coding, or questions that facilitate communication and review about coding
  - (explaining coding nuance) "Do you have questions about that?"
  - "So how was coding this week?"
  - "What did you see as the components included in that comment, that made it a level 2?"
  - "Is coding getting easier?"
  - (live coding, supervisor pauses video) "What behavior did you see?"
  - "What time are you looking at?"
  - "Did you get my coding?"
- **Client**: question relates to a specific family/case
  - "So will dad be there for future sessions?"
  - "What do you think is this mom's strongest target at this point?"
- **Other:** question is too vague to code as one of the above, or addresses other/general ABC-related topics
  - (often at end of supervision session) "Do you have any other questions?"
  - "How was your first session?" "How did your session go?"
  - "How did the pudding session go?"
  - "Did you get your montage together?"

Comment Question/Prompt Specifiers

Q/P comment topic & characteristics	
Q/P topic - comment frequency	24
Q/P topic - comment level/components	25
Q/P topic - comment on/off-target	26
Q/P topic - negative comments	05
Q/P topic - comment other	27
checking for understanding A	28
(SUP) eliciting reflecting/self-supervision	29
platform (summary of other's talk)	30
open question	31
closed question	32
prompt/probe	33

With comment questions/prompts, both topic and type (open/closed/prompt-probe) should always be coded. In addition, a few other specifiers (discussed below) may be coded, but do not have to be coded.

Topics are very similar to other supervisor behaviors related to commenting, and include:
- Comment Frequency
- Comment Level/Components
- On-/Off-Target
- Negative Comments
- Comment Other (including vague/unclear/general topics)

Types include:
- **Open Question** – uses words like "how," "why" and "what"
  - "What did you think of the coding system?"
  - "What questions do you have about that?"
- **Closed Question** – includes yes/no questions, or questions with either/or responses; use words like "Do," "Is" "Are," and "Can."
  - "Do you have any questions about that?"
  - "So are you confused about that, or are you good?"
  - "Does that make sense?"
  - "So you understand why that would be intrusive, right?"
  - "… if that makes sense?" (in up-tone)
- **Prompt/Probe** – a non-question statement that prompts a response
  - "So when you emailed, you said that you felt more confident about coding this week."
  - "This looked like a really chaotic session – it must have felt super tough when you were there."
  - "Tell me more about that."

- If the supervisor asks multiple questions related to the same topic, Open Question trumps Prompt/Probe which trumps Closed Question, if the supervisor uses multiple types.
  - e.g., "How did you feel commenting this week? Did it feel a little easier?" would be coded as an Open Question.

<u>Other Question/Prompt Specifiers</u>:
- **Checking for Understanding** – <u>coded for supervisor, only</u>, when the supervisor is assessing the coach's knowledge, understanding, and following of the conversation
  - "Do you have any questions?" (most common!)
  - "Does that make sense?"
  - "Why did you code that as a 1?"
  - "Do you remember that part of the clip?"
- **Eliciting reflecting/self-supervision** – <u>coded for supervisor, only</u>, when the supervisor asks questions that are intended to elicit parent coach reflection on experience of ABC/coding or self-directed feedback about performance
  - "How did you feel about your session 7?"
  - "Was the coding for this clip easier?"
  - "Are you excited for session 10?"
  - "How did you think you did in terms of the goals we talked about last week?"
- *Note*: Checking for understanding and Eliciting reflecting/self-supervision are usually mutually exclusive.
  - Exception examples:
    - "I don't know if you were trying to put both of them [FTL and nurturance] in there, because it is both, or you were trying to correct yourself after…"
    - "Did you have any questions? Did you run into any problems other than [already discussed coding question]?"
    - "Did you have any questions or concerns about your commenting in this session?"
    - "Do you have any other questions or comments?"
- **Platform** – a platform is when the person reflects back, or rephrases what the other person has said
  - PC: "This session was so stressful! I had a really hard time staying focused on mom's behavior so I could make comments." Supervisor: "It was hard to focus on mom's behavior because there was so much going on. Did you figure out any strategies that helped with that, over the course of the session?"

Coding Question/Prompt Specifiers

For **Coding Questions**, the only difference from Commenting Questions is with Question Topic:

Q/P coding topic & characteristics		
Q/P topic - behavior coding		34
Q/P topic - comment coding		35
checking for understanding b		36
(SUP) eliciting refleciton/self-supervision…		37
platform (summary of other's talk) b		38
open question b		39
closed question b		40
prompt/probe b		41

For **Topic**, you will choose whether the comment is discussing behavior coding or comment coding.

- **Behavior Coding** – discusses coding of parent behaviors
  - o "Does that make sense, that we wouldn't code a 2 because it's a reasonable limit?"
  - o (discussing behavior coding topic) "Do you have any questions about that?"
- **Comment Coding** – discusses coding of coach comments
  - o "When you coded that as a level 2, what components were you seeing?"
  - o (discussing comment coding info) "Does that make sense?"
- **Other/Unclear/General = *don't code either of the above***
  - o Note that this is different from Comment Coding, where "Other" is an option
  - o "Did you get the coding I emailed?" (this is a common one!)
  - o "Did you have any questions when you looked at my coding sheet?"

Client Question/Prompt Specifiers

There are no specific topics for Client Questions:

Q/P characteristics		
checking for understanding		g
eliciting reflection/self-supervision		h
platform (summary of other's talk) c		i
open question c		j
closed question c		k
prompt/probe c		l

"Other" Question/Prompt Specifiers

There are also no specific topics for "Other" Questions. The modifier options are the same as for Client Questions, pictured above.

Informing/Recommending

This category is coded when the supervisor provides information or advice about commenting or coding, in general (i.e., not as feedback about the coach's performance).

Note: when deciding between coding as "feedback" versus "informing/recommending," listen for a "you" referencing the parent coach. If you don't hear it, it's probably informing/recommending.

There are 3 types of Informing/Recommending behavior:

- **Other** – general ABC advice and providing information on the process of ITM supervision; can include discussing the supervisor's perspective and/or lack of expertise in an area, directing coach to ask question of clinical supervisor
    - o "Since this parent has a low IQ, you may want to simplify your language when discussing manual content."
    - o "You know, I'm not sure about that. I think that might be a question for your clinical supervisor."
    - o "So you'll want to bring trash bags, or plastic sheets to put under the parent for the pudding activity, so she's comfortable with getting messy."

    - o Modifiers for Other:

Topic4	
ITM supervision process	6

    - o **"ITM supervision process"** is the only topic we care about capturing; all other topics should be coded without a modifier. This is coded when the supervisor is explaining aspects of ITM supervision.
        - "I just select the clips randomly, using a random number generator online, and then I screen to make sure the parent is on-screen enough to make the clip codeable."
        - "So each week, you'll send me your videos, I'll assign a 5-minute clip to code, and you'll send your coding to me, and then I'll send my coding to you."

- **Commenting** – providing advice or information about commenting
    - o Modifiers for Commenting:

Topic	
Frequency	a
Level	b
On/off-target	c
Negative Comments	02
Other	d
recs are case-specific	07

- o **Frequency** – recommendations, information, or advice about making more frequent comments
    - ▪ "So next week you have session 10. We often see a dip in frequency because you're spending more time watching videos with the parent. So a couple tips for keeping your rate up are to feel free to interrupt the video to make a comment, and then play it over again…"
    - ▪ "One thing you can do to try to keep your comment rate up with this family, even before you start making more negative comments, is to focus on commenting on delight."
- o **Level** – recommendations or advice about different components of comments
    - ▪ "For nurturance and following the lead, your behavior descriptions have to have a child piece and a parent piece."
    - ▪ "The outcomes sheet can be pretty overwhelming, so I recommend just memorizing one outcome per behavior target at first."
- o **On/Off-Target** – recommendations or advice about avoiding off-target comments
    - ▪ "For a parent who switches back and forth between following and not following like this, it's easy to be off-target if you just make your comment a little too late, after they switch to not following. But as long as you label the behavior clearly, you can still be on-target."
    - ▪ "One reason we want to avoid child-directed comments is that parents are less likely to listen to the comment, if they think we're just talking to the child."
- o **Negative Comments** – recommendations or information about scaffolding, shaping and negative comments; "you could have" comments
    - ▪ "Shaping comments are the least threatening of the ways you can respond to negative behavior."
    - ▪ "So you always want to be monitoring the parent's reaction to make sure they're not reacting negatively to the comment."
    - ▪ "And as soon as possible, after you make a negative comment, you want to be looking for a positive behavior to

119

comment on – even if it's the behavior you just suggested the parent do!"
- "So you want to keep the ratio of positive to negative comments about 4 to 1."
- o **Other** – recommendations about commenting that do not fit in the above categories, including commenting on specific behavior targets
  - "In session 1, you should start commenting on all the targets, not just nurturance. You can say something like 'we'll be talking about that more in a few weeks.'"
  - "Meta-comments talk generally, rather than specifically, about the parent's behavior."
- o ***Recs are case specific*** - this code should be coded in addition to one of the above when the supervisor directly references one of the parent coach's current cases.
  - "One thing you can do to try to keep your comment rate up with this family, even before you start making more negative comments, is to focus on commenting on delight."
  - If the coach discusses a commenting strategy specific to a particular session, but doesn't personalize these recommendations to the family, do NOT code as case-specific
    - "So in session 3, you can start making scaffolding comments, but you want to keep it only to when you're having the family do the activities."

- **Coding** – providing advice, information, or recommendations about coding
  - o Modifiers for Coding

Topic2		
Behavior coding		e
Comment coding		f

  - o **Behavior Coding** – discussing rules, guidelines, and rationale/value related to describing and coding parent behaviors; also guidelines for when not to code behaviors
    - "Nurturance can be either when the child is distressed, or when the child is looking for affection."
    - "It's important to include the time each behavior occurs, so that when we compare our sheets, it's easier to figure out if we were seeing the same behavior."
    - "When the child reaches for the parent to pick them up, that can be tough to figure out if it's nurturance or following the lead. If the child snuggles in after they are picked up, and just seem like they want to be with the parent, that would be nurturance. But if they seem to want to reach something, see the laptop better, or just use the parent as a seat, that would be following the lead."

- "So things like keeping the child from standing on the couch, sharing toys with siblings, and eating their snack while sitting down instead of wandering around the room, those are all reasonable limits, so we wouldn't code them as not following."
- "You don't need to write that the parent is talking to you. All you need to code is the stuff that is an ABC target."
- "If you scroll over here, there's a table summarizing the parent behaviors for the clip, and you can use it to track progress."
  - **Comment Coding** – discussing rules, guidelines, and rationale/value related to describing and coding comments; also guidelines for when not to code comments
    - "When there's a pause, and a new example of the same behavior occurs, we would split it up into 2 different comments."
    - "That would be a level 2 because it has both the full behavior description and the outcome."
    - "You should transcribe exactly what you say. I usually pause and then listen to the comment a second time."
    - "So when you make a scaffolding or negative comment, you want to code the comment target as the behavior you are responding to – the negative behavior – not the positive behavior that you're trying to encourage.
    - "So this table is the first thing you want to look at when you finish your coding, so you can evaluate how you're doing. The numbers you want to look at are…"

Client Discussion

This behavior is coded for both supervisors and coaches when they are discussing individual cases/families (generalizations about families/cases the coach is seeing will probably be coded under reflection or irrelevant ABC/case-related talk).

Modifiers:

Client topic		
topic - case conceptualization		m
topic - outcomes/progress (TRUMPS con...		n
topic - other		o
specificity - specific client behavior		p
specificity - client behavior 5 min clip		q
specificity - client more general		r
client valence - positive		s
client valence - negative		t
client valence - neutral		u
link to ITM		v

Topic Modifiers
- **Case Conceptualization** – this is the most frequently coded topic for client discussion. This includes discussion of client behaviors during session, and other information relevant to understanding the family and/or tailoring ABC to them (e.g., risk factors, history, attitudes toward ABC, future plans for foster placement/reunification, etc.)
    - o "This mom shared with me that she is depressed."
    - o "So that's actually stepdad you're seeing in the video. The child's biological dad is not involved."
    - o "Even though she's really intrusive with the baby, she does delight a lot."
    - o "There was a nice opportunity for nurturance in this clip, when the baby started fussing, and we got to see how mom responded."
- **Outcomes/Progress** – this is less frequent; when statements include any discussion of client progress, change, or outcomes, this should be coded (trumps case conceptualization)
    - o "Mom really seems to be understanding following the lead now."
    - o "It was so cool to see her catch herself and talk to you about how she would have yelled in the past, but was trying to handle it differently."
    - o "Looking at the coding sheet, I see so many more 1's in this session than in the first couple sessions."
- **Other** – client discussion that seems relevant but does not fit in the first 2 topic areas; for example, discussion about recruitment, anticipation/wondering about what will occur in future with this client
    - o "I'm interested to see whether she'll start responding to the comments more, as you get farther into the manual."
    - o Note: irrelevant discussion about clients is coded under **Other**(*participant*)/ **Irrelevant ABC-/case-related talk**

Specificity Modifiers

- **Specific Client Behavior** – discussion of one or a handful of specific client behaviors
  - "I noticed that she tickled the baby a few times in this clip."
  - "When the baby started coughing, it was really sweet how mom nurtured her."
- **Client Behavior 5 Min Clip** – discussion of behavior in coded clip (may be more or less than 5 minutes, e.g., if a play video is coded)
  - "Looking at this coding sheet, there's a lot of back-and-forth between 1's and 2's."
  - "There weren't many parent behaviors in this clip, which makes it hard to comment."
    - Note that this could be coded as part of comment feedback, rather than client behavior, if preceded by evaluation of coach's commenting (e.g., "You made two comments in this clip, but…")
- **Client More General** – discussion of client behavior more generally
  - "She's gotten so much better at following."
  - "This mom is pretty teachy."
  - "Dad really enjoys talking to me. Which is good, but it also means that he ignores the baby a lot."

Valence Modifiers
- **Positive** – when the tone/content of what is said about the client is warm/positive
  - "I've seen them change so much, it's amazing."
- **Negative** – when the tone and content of what is said about the client is negative, complaining, overwhelmed, or disappointed
  - "I really feel like I can't get through to her. She isn't listening to what I say."
  - "There was SO much tickling in the play session, it was really hard to watch and not say anything."
- **Neutral** – when the tone of what is said about the client is neutral or objective
  - Neutral comments about clients may include negative evaluations, but have an objective, dispassionate tone
  - "He's a tickler, so that'll be something to work on."
  - "Even though she's really intrusive with the baby, she does delight a lot."
- When coding Client Progress, valence captures the direction of the progress, more than the tone

Link to ITM
- This modifier is coded, in addition to the 3 categories above, when ITM commenting is referenced when discussing clients.
  - "There weren't many parent behaviors in this clip, which makes it hard to comment."
  - "She seems to be pretty open to your suggesting comments."

- o "I've really seen big changes in following the lead in the past 2 sessions, and I'm sure it's because you've been making so many comments."
  - Note: When deciding between "Client Discussion" with "Link to ITM Specifier," and "Comment Feedback" with "Ref – Client," consider the subject/topic of the sentence. Here, the topic is the client (though it is implied). If the subject were the coach, instead (e.g., "You've been making so many comments with this family, and I can really see how it's been helping them to follow the lead more"), it should be coded as "Comment Feedback."

Active Learning

Two supervisor-led active learning behaviors are coded (Modeling, Live Coding). Role-Play is also listed, but is always coded as a parent coach behavior. Additionally, Thinking/Silence, is listed, but is coded under "Other" participant (described above under "Other" participant section).

The 2 supervisor-led active learning behaviors are:
- **Modeling** – coded when <u>supervisor</u> provides example of language that could be used in session, usually an example of an "in the moment" comment that could be made, but can also include language for explaining the idea of in the moment commenting to the parent, explaining that ABC is all about the parent-child interaction, etc.
  - o **EACH IN THE MOMENT COMMENT SHOULD BE CODED AS A SEPARATE "MODELING" BEHAVIOR.** This means that you should create a small pause between each example of comment (e.g., if supervisor says, "You could say, 'Oh, she's got the donkey toy' or 'I'm sorry, I'm keeping you wrapped up in this conversation and not letting you respond to her,'" you would code 2 separate Modeling behaviors, separated by a few milliseconds.
  - o <u>If the supervisor's "modeled" comments are being read off a worksheet,</u>
    - <u>Code "modeling" if the parent coach CANNOT see the worksheet (i.e., no shared screen, or reference to having the coach open a file on her end)</u>
    - <u>Do not code modeling (likely, just "Informing/Recommending – Commenting") if the parent coach CAN see the worksheet</u>
      - <u>But if the supervisor gets creative with their reading and deviates from reading word-for-word, code it as a modeled comment</u>
  - o <u>Code modeling even when supervisors are providing examples of bad or off-target comments</u>

- **Live Coding** – coded when either parent coach or supervisor explains coding, talks through thoughts on coding, or discusses behavior, based on video example or worksheet viewed in session.
  - o There will often be pauses in speech during live coding while the video plays. Don't bother switching to coding "Thinking/Silence" for pauses shorter than 5-10 seconds (this is a pretty big range; the high end of 10 seconds would apply if the video continues playing and there is a reinitiation of live coding talk, whereas if the video ends or is paused after ~8-10 seconds of silence, you could code the final 8-10 seconds as Thinking/Silence).

Active Learning: Realism Specifiers

◢ Realism		
Video continuous		w
Video paused		x
PC's coding sheet		y
Worksheet		z
Other (e.g., spontaneous)		01

- **Video continuous** – this is coded when the video is playing during active learning activities; common combinations include:
  - o Other – Thinking/Silence – Video continuous (supervisor & coach watching video silently)
  - o Supervisor – Live Coding – Video continuous (supervisor talking through coding as video plays)
  - o Coach – Role Play – Video continuous (coach making comments as video plays)
- **Video paused** – this is coded when the video is paused during active learning activities; common combinations include:
  - o Supervisor – Live Coding – Video paused (supervisor pauses video and then talks through coding)
  - o Supervisor – Modeling – Video paused
  - o Coach – Role Play – Video paused
- **PC's coding sheet** – this is coded when the active learning activity is based off of a coding sheet from a recent session of the parent coach; common combinations include:
  - o Supervisor – Modeling – PC's coding sheet
  - o Coach – Role Play – PC's coding sheet
  - o *Note*: explaining coding done prior to session, without video, should be coded under "Coding Feedback" or "Informing/Recommending"; thus, Live Coding should never be paired with "PC's coding sheet"
- **Worksheet** – this is coded when the active learning activity is based off of a worksheet (i.e., coding sheet developed/introduced for didactic purposes, not

from real sessions of the coach); would also include "practice coding" sheets of sessions from ABC website; common examples include:
- o Supervisor – Modeling – Worksheet
- o Coach – Live Coding – Worksheet
- **Other (e.g., spontaneous)** – this is coded when the active learning activity does not fit into the above Realism categories; most often, it is coded when Modeling or Role Play is integrated into other statements; common examples include
  - o Supervisor: Informing/Recommending – Commenting AND Modeling – Other (spontaneous)
    - § "So one way you can start commenting about following the lead even in session one is to just make a comment, like, 'He handed you the toy and you took it, that's following the lead. We'll talk more about following the lead in a couple weeks.'"
    - § "Since you feel like this parent is more sensitive to negative feedback, you could use more spotlighting the good comments, which focus on the positive pieces of an overall negative behavior, like if she's tickling him and then stops, you could say, "Good job noticing that he was getting a little overwhelmed there, and slowing down the interaction."
  - o *Note*: when Modeling or Role Play occurs within another activity (e.g., Informing/Recommending, Goal Setting, Reflection), code it over/on top of the other activity. This is the only time you'll code 2 different activities for the same participant at the same time. However, only code the speech that models what to say, not the other speech (i.e., the parts in 'quotation marks' above) – so Modeling/Role Play will just overlap for a portion of the other activity.

Goal Setting

Two types of goal setting are coded, goals about commenting and goals about coding.

Goal Setting – Commenting is coded when supervisors suggest, or coaches make, goals for the coaches' upcoming performance with regard to commenting. This category can be difficult to distinguish from Informing/Recommending, and should only be coded when it is fairly explicit that this statement represents goal setting – communicating a *target* or *focus* for improvement for the coach's performance in future sessions, as opposed to general recommendations. For supervisors, this seems to occur most often in the final minutes of a supervision session, as supervisors are recapping and highlighting the biggest things to focus on in future sessions.

Modifiers for Goal Setting – Commenting:

Goal topic - comments	
goal - comment frequency	42
goal - comment level/components	43
goal - comment on/off-target	44
goal - negative comments	06
goal - comment other	45
specificity - 1 (very)	46
specificity - 2 (moderate)	47
specificity - 3 (low)	48

- **Topics**: Topics for goals about commenting are the same topics described above for other aspects of supervision, such as feedback
    - **Comment Frequency**
    - **Comment Level/Components**
    - **Comment On/Off-Target**
    - **Negative Comments**
    - **Comment Other**
- **Specificity** modifiers range from 1 (most specific) to 3 (least specific)
    - **1:** gives a numeric goal
        - "For next week, I want you to try to make one more comment per 5 minutes."
    - **2:** gives a less specific goal, but provides at least one specific strategy to meet the goal or seems more thoughtful/process-oriented than a level 3 goal; can include more specific goals (strategies), with unstated overarching goal
        - "In your next session with the family, I want you to try to raise your level of comments, and to do this by adding target labels into your comments."
        - "I want to be more in-tune to when she's following vs. not following, and about when to use the [following] label or not to" (i.e., overarching goal is not to be off-target with FTL comments)
    - **3:** gives a non-specific goal without strategies
        - "For your next session, just include as many outcomes as you can!"
        - "So for next week, I want you to focus on using those negative comments with this family."
        - "In your pudding activity, comment as frequently as you can!"

Goal Setting – Coding is coded when the supervisor suggests, or the coach describes, goals for the coaches' upcoming performance with regard to coding. Again, this category can be difficult to distinguish from Informing/Recommending, and should only be coded when it is fairly explicit that this statement represents goal

setting – communicating a *target* or *focus* for improvement for the coach's performance in future coding assignments, as opposed to general recommendations.

Modifiers for Goal Setting – Coding:

Goal topic - coding	
goal - behavior coding	49
goal - comment coding	50
specificity - 1 (high) b	51
specificity - 2 (moderate) b	52
specificity - 3 (low) b	53

- **Topics**: Topics for goals about coding are the same topics described above for other aspects of supervision, such as feedback
    - **Behavior Coding**
    - **Comment Coding**
- **Specificity** modifiers range from 1 (most specific) to 3 (least specific)
    - **1:** gives a numeric goal
        - "For next week, I want you to try to code at least 3 behaviors that you didn't comment on in the clip."
    - **2:** gives a less specific goal, but provides at least one specific strategy to meet the goal
        - "For next week, I want you to focus on describing the behavior clearly in your coding, so that it's easier to compare what we're each coding. So a strategy for doing that is to try to describe the behavior like you would in your comments – Child does this, Parent does that."
        - "When you're coding your next assignment, focus on catching more of the parent's delight. One strategy that I think is helpful is to watch the clip another time, after you finish your coding, just focusing on finding delight that you missed the first time."
    - **3:** gives a non-specific goal without strategies
        - "When you're coding your next assignment, focus on catching more of the parent's delight."
        - "So for next week, try to code more behaviors that you didn't comment on in the session."

Supporting/Understanding

Supporting/Understanding is a behavior that is coded only for <u>supervisors</u>, in which supervisors make brief (usually non-specific) encouraging statements, or communicate understanding/acceptance of what the parent coach has said.
- <u>Only statements of >3 words should be coded</u>; statements shorter than this ("Mhmm," "Yeah," "Totally," etc.), which usually occur as the coach is speaking, are too much work to code.

- "Everyone feels overwhelmed in their first ABC session!"
- "Hang in there, it's really tough right now, but it's going to feel so much easier with your next case!"
- "You're doing awesome – keep up the great work!"
- "Good luck in your next session!"
- "That sounds really stressful."

Other – Discussion of PC's Workload

Discussion of PC's workload is coded when the supervisor engages in discussion of the parent coach's caseload of ABC clients, time spent on ABC-related activities (e.g., recruiting families, editing video, etc.), balance of ABC with other job responsibilities, etc. Include questions about workload under this category, rather than under Questions. However, questions/discussion that seems to be geared more toward planning (e.g., "So what cases will you be seeing this week?" "When do you think you'll pick up your next case?") should be coded under Other – Planning/Scheduling/Agenda Setting. This code is about capturing coach's reflection/worry/rumination about an overwhelming schedule, and supervisors' support for the coach and desire to understand the coach's other job responsibilities that can make learning ABC harder.

Parent Coach Behaviors

When you hit "p" for parent coach, the following behaviors will come up:

Provision of feedback			
Comment Feedback		f	F
Coding Feedback		e	E
Questions/Prompts			
Q/P topic - comment		q	Q
Q/P topic - coding		a	A
Q/P topic - client		x	X
Q/P topic - other		o	O
Client Discussion			
Client discussion		d	D
Active Learning			
AL Modeling		m	M
AL Role-play		r	R
AL Thinking/silence *cod…		t	T
AL Live coding		l	L
Goal Setting			
Goal topic - comment		j	J
Goal topic - coding		u	U
Response to input/qu…			
Responding to input/qu…		z	Z
Reflecting			
Reflection		1	1
Other (participant-spe…			
Discussion of PC's workl…	Parent Coach	w	W

Many of the behaviors coded for parent coaches are the same behaviors, with the same modifiers, as the behaviors for supervisors. Refer to earlier sections of the manual for guidelines and examples on how to code these. There are only a few new/different behaviors, discussed later.

Supervisor-only behaviors, NOT coded for parent coaches, are:
- Informing/Recommending
- Supporting/Understanding

Additionally, several sub-behaviors and modifiers should NOT be coded for parent coaches, but unfortunately show up in the coding scheme because of the coding program:
- Active Learning – Modeling (for coaches, Role-Play replaces Modeling)
- Active Learning – Thinking/Silence (always coded with "Other" as participant)
- Question/Prompt Specifiers
  - Checking for understanding
  - Eliciting reflection/self-supervision

Examples of Parent Coach Behaviors Previously Described as Supervisory Behaviors

Feedback
When feedback is coded as a coach behavior, it always represents "self-feedback."
- **Comment Feedback**
  - "I missed so many opportunities to comment in this clip!" (coded as Comment Feedback with modifiers: topic – frequency, valence – negative, reference – other/unclear, specificity – 5 min. clip)
  - "I'd give myself a B this week. I felt like I commented a lot, but I kept forgetting to include the child behavior in my descriptions." (coded as 2 different instances of Comment Feedback; 1st two sentences coded with modifiers: topic – frequency, valence – neutral, reference – other/unclear, specificity – more general; 3rd sentence coded with modifiers: topic – level/components, valence – neutral, reference – other/unclear, specificity – more general)
  - For reference, comment feedback specifiers are:

Comment Feedback		
Topic - frequency		L
Topic - level/components		M
Topic - on/off-target		N
Topic - negative comments		03
Topic - other		O
Valence - positive		P
Valence - negative		Q
Valence - neutral		R
Reference - norms/other PCs		S
Reference - certification criteria		T
Reference - PC's past performance		U
Reference - goals		V
Reference - ITM supervisor		W
Reference - developmental		X
Reference - just #s		10
Reference - client		04
Reference - other/unclear		Y
Specificity - single comment		Z
Specificty - in 5-min clip		00
Specificity - more general		7

- **Coding Feedback**
  - "Looking at your sheet, I realized that I coded a lot of stuff that I didn't have to." (coded without topic, because unclear; valence – neutral; ref – ITM supervisor, specificity – 5 min)
  - "I missed so many of the behaviors you coded!" (coded as topic – behavior; valence – neutral/negative (depends on PC's tone); ref – ITM supervisor, specificity – 5 min)
  - For reference, coding feedback specifiers are:

Coding Feedback	
Topic - behavior coding	11
Topic - comment coding	12
Valence - positive b	13
Valence - negative b	14
Valence - neutral b	15
Ref - norms/other PCs	16
Ref - certification criteria	17
Ref - PC past performance	18
Ref - goals	19
Ref - ITM supervisor	20
Ref - developmental	21
Ref - other/general	22
Specificity - single behavior	23
Specificity - in 5 min b	8
Specificity - more general b	9

Questions/Prompts

Question Topics
- **Comment:** question relates to understanding of comment topic
    - "So I can hold off on making any negative comments until session 3…"
    - "Can you explain the different outcomes again?"
- **Coding:** question relates to understanding of coding topic
    - "Why was that a 2?"
    - "So for every 5-10 seconds, I would group the behaviors together?"
    - "What if … (hypothetical behavior scenario)?"
- **Client**: question relates to a specific family/case
    - "What did you think of the play video I sent you for my newest case?"
    - "Have you seen cases where a mom and dad are this different in their behaviors before?"
- **Other:** question is too vague to code as one of the above, or addresses other/general ABC-related topics
    - "What is the password for the website?"
    - "How long should I plan to spend on making the montage?"
    - "What kinds of toys are best to bring?"

Other Question Specifiers
Other than question topics, the only modifiers coded for parent coach questions are:

- **Platform** (coded if the coach repeats back part of what the supervisor has said; this often happens when the coach seems to be checking that they understood what the supervisor said)
- **Open Question**
- **Closed Question**
- **Prompt/Probe**
- As with supervisor question, <u>one, and only one</u>, of the final three listed modifiers should always be coded for any given question/prompt.

Client Discussion

See Supervisor section for examples; many of the examples provided there are information provided by the coach.

Active Learning – Live Coding

See Supervisor section for description.

Goal Setting
- **Goal Setting – Commenting**
    - "Okay, this week my goal is to comment on as much delight as I can."
    - "For next session, I'm going to try to not worry about interrupting the parent, to just say the comment."
- **Goal Setting – Coding**
    - "This week, I'm going to set aside a full hour to do my coding, and make sure it's not the night before we meet."
    - "So this week, I'll try to just include the ABC behaviors, and leave out the extra stuff."

<u>Additional Parent Coach Behaviors</u>
(behaviors not coded for supervisors)

<u>Active Learning</u>

- **Role Play** – coded when <u>parent coach</u> rehearses, practices, or provides example of language that she could use in session <u>in the future</u> (including both comments and other ABC session language)
    - If the coach is reporting about comments she made, or what she said, in a <u>past</u> session, do <u>not</u> code Role Play (most likely will code Reflection)
    - **EACH IN THE MOMENT COMMENT SHOULD BE CODED AS A SEPARATE "ROLE PLAY" BEHAVIOR.** Stop coding for a brief time (a few milliseconds) so that the number of role played comments can be captured from the coding.
    - Role play may overlap with another parent coach behavior, such as a Reflection.

- ▪ "I think I should have been more specific, like, 'Dad, when he started to cry, you picked him up and held him.'"
- ▪ (prompted by supervisor in live commenting activity) "Hmm… I guess I could say 'Good following when he handed you the bubbles.' But what's tricky for me is that mom took them but then put them down to the side."

Reflection

A reflection is a statement the coach makes about their experience of doing ABC, that is not specifically evaluatory enough to be coded as self-feedback. Reflections include a coach's report on what she did in session or when coding, how she felt, what she saw/observed, and what she thinks. Discussion of what happened in session that is exclusively focused on family behaviors will usually be coded as Client Discussion.

Three different aspects of reflections should be coded as Specifiers

◢ Topic3		
1- coding		A
1- commenting		B
1- general ABC		C
2- affective		D
2- descriptive		E
3- spontaneous		F
3- elicited by supervisor		G

1 – Topic: the topic of the reflection can be coding, commenting, or general ABC.
- • **Coding**
  - o "I guess I didn't know I could code that."
  - o "I think what happened was that I was trying to cut and paste. Because I went back through and saw things I had missed, and then put them in at the end, but then I was trying to put them in the right order…"
  - o "That was something I was really confused about when I was coding."
- • **Commenting**
  - o "I feel like this parent has so many positive behaviors that I get caught up in praising her, and forget to use the negative comments."
  - o (long pause when coach is trying to think of comment she could make, in commenting activity) "I'm doing it again, where I get stuck thinking about what I want to say."
- • **General ABC**

- o "So far, all the parents have loved ABC. One parent actually thanked me and said she couldn't wait until next week."
- o "I feel more nervous about session 2 than I did about session 1, I think because of the nurturing thing."

2 – Affective vs. Descriptive: this assesses whether the reflection discusses feelings/preferences or just thoughts/observations
- **Affective** – reflection discusses the feelings experienced by the coach, including feeling nervous, confused, or excited; or expresses preferences (i.e., liking or disliking). Reflections with any affective statements will be coded here, even if they also include some description (i.e., affective trumps descriptive). If a coach's reflective statement does not include affective words, but the tone of voice clearly communicates the coach's feelings, it may also be coded as an affective reflection.
  - o "So the beginning wasn't following the lead, but then it was, so I was confused about that."
  - o "I like the coding, I think because I'm such a detail-oriented person."
  - o (supervisor says coach made an amazing shaping comment) "I want to hear more about this amazing shaping comment, because [dark/sarcastic tone] it didn't feel like it"
  - o other affective reflection keywords: comfortable, uncomfortable, awkward
- **Descriptive** – reflection objectively discusses coach's experience, thoughts or observations
  - o "My co-worker who is also doing ABC told me I had to transcribe everything, so that's what I did."
  - o "I was trying to mix up my comments, like not do all level 3's."
  - o "I really need to work on my scaffolding and shaping comments, I've talked about it with [Clinical Supervisor]."

3 – Spontaneous vs. Elicited By Supervisor: this assesses whether the reflection is preceded by a supervisory effort to elicit reflection, or occurs spontaneously.
- **Spontaneous** – may be preceded by
  - o Supervisor feedback
  - o Supervisor informing/recommending
  - o Supervisor supporting/understanding
  - o Coach talking about something else
  - o Etc.
- **Elicited By Supervisor** – reflection is preceded by a Question/Prompt from supervisor
  - o Note that the supervisor's question may not itself be intended to elicit a reflection (e.g., "Does that make sense?"). In this case, the Question/Prompt should not be coded with the "eliciting reflection/self-supervision" specifier, but the Reflection should be coded as "Elicited by Supervisor."
  - o You may code multiple reflections, separated by other coach behaviors (e.g., discussion of workload, self-feedback, etc.) or

supervisor behaviors (e.g., supporting/understanding). If the coach continues with the same line of reflection, second instances can still be coded as "elicited by supervisor," even if they are not directly preceded by a supervisor's Question/Prompt. However, if subsequent reflections move on to other topics (e.g., switch from answering supervisor's question about how commenting felt, to "But what was really hard for me was coding! I had to keep pausing and rewinding the video!"), they should no longer be coded as "elicited by supervisor."

Responding to Input/Question

Responding to Input/Question includes coach responses of >3 words, which do not fit other categories, particularly Reflections. This category captures coaches' responses to supervisors' informing/recommending, feedback, goal-setting, and questions that don't elicit other types of coach behaviors.

There are four specifiers for Responding to Input/Question:

◢ Responding to Input		
valence - positive		H
valence - negative		I
valence - neutral		J
elaboration		K

Valence Specifiers

Positive – positively valenced responses are coded primarily based on tone of voice and facial affect. In some cases, content of the response may be enough to code a positively valenced response; in these cases, the content should use superlatives (e.g., great, wonderful, exactly) and go beyond just expressing agreement or understanding.
- "I get it now. That is so helpful."
- "Yes, (chuckle) that's what was going on!"
- "So I'll keep that in mind, definitely, for our next session."

Neutral – for most coaches, the majority of responses will have a neutral valence. These often express agreement, or add information, but are not accompanied by strong affect or language.
- "Oh, okay, that makes sense."
- "Yeah, I can do that next time."
- "Nope, that sounds good."
- (in response to "Do you have any questions?") "I don't think so."
- "Okay, we'll see how that goes."

Negative – negatively valenced responses express disagreement or defensiveness. This can be communicated by tone of voice, interrupting the supervisor, or language.
- (following supervisor's explanation of why one of coach's comments was a "meta-comment") "When I said 'all the kids' I meant her 4 children, not like all the children in the world."
- (following supervisor's explanation of why she didn't code a behavior because she considered it appropriate limit-setting) coach interrupts supervisor: "Well it wasn't so much that, it was that they kept telling him to stop."

Elaboration

Elaboration is coded as a Response to Input/Question specifier when the coach's response goes beyond expressing agreement or understanding, and adds novel information.
- Both negative responses above would also be coded with the "elaboration" specifier
- (responding to question about where the coach sets up her camera) "So they have a front room where they meet, and the kitchen is behind it. So I put it in the corner so I can try to get the kitchen doorway, because the other kids are going in and out of the kitchen."

Other Coding Details

Stuff that is Not Coded

Do not code brief (1-3 word) statements expressing agreement/understanding.

Do not code off-topic chat ("How are you?" "We got a big snowstorm here" etc.)

Do not code silent periods when either party is looking for the coding sheets in email, etc.

Do not code anything (including Tech Issues) before the coach is online (unless it is clear that the coach is trying to communicate with the supervisor, e.g., the supervisor is talking on the phone to the coach, or the coach's face is visible but sound is not working).

Do not code apologizing/discussing miscommunications about timing (e.g., "I'm sorry, I thought we were going to meet at 3.")

Do not code discussion of sending videos (e.g., which videos need to be sent) that seems to be for *research*, rather than supervision purposes (e.g., sessions that were conducted prior to the session being reviewed in the current supervision session, sessions that EB says are missing).

- Discussion of sending videos for supervision purposes will be coded under agenda setting/planning.

Beginning and Ending Recordings

Begin recording (hit the red dot) at the beginning of the video, even if there is a long period of the supervisor waiting for the coach to join the supervision session. (We'll take this out when analyzing data, it's important for us all to start coding at the same exact time so reliability analyses match second-to-second.)

End recording when the coach leaves the call (indicated by Skype noise, face going off-screen, etc.). It's more okay for reliability analyses if we don't end the recording at the same exact second.