ORIGINAL ARTICLE

Effects of Fidelity-Focused Consultation on Clinicians' Implementation: An Exploratory Multiple Baseline Design

EB Caron[1] · Mary Dozier[2]

Abstract

Identification of effective consultation models could inform implementation efforts. This study examined the effects of a fidelity-focused consultation model among community-based clinicians implementing Attachment and Biobehavioral Catch-up. Fidelity data from 1217 sessions from 7 clinicians were examined in a multiple baseline design. In fidelity-focused consultation, clinicians received feedback from consultants' fidelity coding, and also coded their own fidelity. Clinicians' fidelity increased after fidelity-focused consultation began, but did not increase during other training periods. Fidelity was sustained for 30 months after consultation ended. Findings suggest that consultation procedures involving fidelity coding feedback and self-monitoring of fidelity may promote implementation outcomes.

Keywords Implementation · Fidelity · Consultation · Self-monitoring

Introduction

Usual care in community mental health settings frequently fails to improve children's functioning (e.g., Warren et al. 2010). Although many interventions have proven effective in clinical trials, usual care clinicians often do not use these interventions (National Advisory Mental Health Council 2001). In addition, evidence-based interventions frequently underperform in the community, as compared to the lab, a discrepancy attributed to lowered treatment fidelity, among other factors (Hulleman and Cordray 2009). The field of implementation science has grown in response to these challenges (Fixsen et al. 2005).

Implementation frameworks can structure the understanding of and response to implementation challenges (Tabak et al. 2012). In Fixsen et al. (2005)'s framework,

core intervention practices are taught to clinicians through communication link processes, including training and consultation. Fidelity is measured to provide a feedback loop between clinicians and those supporting them. Accurately identifying the core components of an intervention is critical, as this defines the focus for training, consultation, and fidelity evaluation. Fixsen et al. (2005) also describe stages of implementation, including exploration and adoption, initial implementation, full operation, and long-term sustainability. At each stage, understanding active ingredients of implementation, or the processes that lead to clinician and agency behavior change, can help close the gap between lab-based treatment trials and the treatments available in communities (Fixsen et al. 2005). The current study focuses on one such active ingredient, consultation, and the fidelity feedback loops that inform consultation.

Consultation, which begins after training and continues while clinicians begin implementing a practice, is critical in promoting clinician behavior change (Edmunds et al. 2013). In addition to dose–response associations between consultation and implementation fidelity (Beidas et al. 2012), several randomized controlled trials of consultation support this idea. Training plus consultation, as compared with training only, has improved clinicians' fidelity to motivational interviewing (MI; Schwalbe et al. 2014), the Incredible Years Parenting Program (Webster-Stratton et al. 2014), and measurement-based care (Lyon et al. 2017).

✉ EB Caron
ecaron@uchc.edu

1 Department of Psychiatry, University of Connecticut Health Center, 65 Kane St., Room 2031, West Hartford, CT 06119, USA

2 Department of Psychological and Brain Sciences, University of Delaware, Newark, USA

However, other studies of consultation have not supported its role as an active ingredient in promoting fidelity (e.g., Darnell et al. 2016). These discrepant findings, along with studies comparing different types of consultation (e.g., Funderburk et al. 2015) and different aspects of the same consultation protocol (e.g., Schoenwald et al. 2004) suggest that different models of consultation are not equal. Such studies raise questions about active ingredients within consultation and the specific processes that make consultation effective (Nadeem et al. 2013).

One potential active ingredient of consultation is feedback (Jamtvedt et al. 2006). Feedback based on fidelity coding is considered the gold standard (Dunn et al. 2016), and has been examined in several studies. Miller et al. (2004) found that written fidelity feedback on three MI sessions was as effective as six consultation sessions in maintaining clinicians' fidelity. Eiraldi et al. (2018) compared Coping Power consultation that included fidelity coding feedback with a basic consultation condition, and found higher fidelity among clinicians who received fidelity feedback. Finally, Weck et al. (2017) found that consultation plus written feedback using the Cognitive Therapy Scale led to larger increases in fidelity to cognitive behavioral therapy than consultation as usual.

Self-monitoring offers clinicians another form of feedback, but is relatively rare in consultation procedures. Studies of self-monitoring have had mixed outcomes in changing clinician behavior, leading to conclusions that clinicians may need additional training in self-monitoring (Dennin and Ellis 2003). Supplementing self-ratings with expert ratings can provide such training. Isenhart et al. (2014) found that two clinicians who coded their own and others' MI sessions in expert-led group consultation showed increasing fidelity over time. They attributed these results to both the feedback component of consultation and the process of learning to code fidelity. Specifically, they theorized that coding sessions would sensitize clinicians to specific aspects of fidelity, allow them to quickly and accurately process session content, and provide them with opportunities to discuss and practice high fidelity responses, which would generalize to in-session behavior (Isenhart et al. 2014). Consultation that integrates self- and expert-rated performance coding has also been used with teaching faculty; Garcia et al. (2017) found that although participants rated self-observation of video more helpful than expert feedback, the student evaluations of previously low-rated teachers improved the following year.

To summarize, understanding how to implement evidence-based interventions in community settings is a critical target for the field. Consultation is considered an active ingredient of implementation, but has not always been found to promote fidelity. Empirical testing of the impact of consultation models is needed. Consultation that includes fidelity feedback has shown promise in enhancing implementation fidelity. Self-monitoring with concurrent expert feedback has been used rarely, and is 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., using a threatening voice tone or actions). The intervention has changed parent behavior in both lab (Bernard et al. 2015b; Yarger et al. 2019) and community settings (Caron et al. 2016; Roben et al. 2017), and has also led to several positive outcomes for children. Specifically, compared with a control intervention (Developmental Education for Families), ABC improved rates of secure and organized attachment (Bernard et al. 2012). ABC has also improved children's executive functioning (e.g., Lind et al. 2017) and diurnal regulation of the stress hormone cortisol (e.g., Bernard et al. 2015a), compared with children in the control intervention.

According to Fixsen et al. (2005), identifying a program's core components is a critical step in implementation. The core component of ABC is "in-the-moment commenting," a process in which clinicians provide live feedback to parents about their behavior (Caron et al. 2018). For example, a clinician might say, "That's a perfect example of nurturance. He whimpered and you 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 predict parent behavior change (Caron et al. 2018). Frequent commenting also reduces likelihood of early dropout from ABC (Caron et al. 2018). As recommended by Fixsen et al. (2005), the ABC fidelity measure is focused on measuring this core component.

After the ABC fidelity measure was developed, clinicians in the lab began to use it for self-monitoring. In a single subject A/A + B design, one clinician received standard lab-based group supervision for 6 months, and then continued standard supervision, with the addition of self-monitoring by coding one ABC session each week, for another 6 months (Meade et al. 2014). Before beginning self-monitoring, the clinician's comment frequency was unchanging; after, however, her comment frequency increased over time (Meade et al. 2014).

These results encouraged the development of consultation procedures for implementation sites that incorporated both

self-monitoring and expert feedback using the ABC fidelity measure. The present study examined the effects of this new consultation procedure in a multiple baseline design with community-based clinicians. We hypothesized that clinicians' fidelity would increase after, but not before, fidelity-focused consultation began. Guided by Fixsen et al.'s (2005) stages of implementation, fidelity was monitored for up to 30 months following cessation of consultation, as we hypothesized that fidelity-focused consultation would promote sustainment of fidelity.

Method

Participants

Participants included seven clinicians at different agencies in Hawaii. Clinicians were recruited for training by a stakeholder at the non-profit Consuelo Foundation, as part of an effort to support the emotional needs of infants involved with Hawaii's Child Welfare System (CWS). Clinicians at agencies with Healthy Start programs, which offered home visiting services to young children in Hawaii's CWS, were prioritized, but clinicians were also recruited from several other agencies, such as a residential home serving pregnant and new mothers. Clinicians were selected through conversations between the Consuelo Foundation stakeholder and agency directors, focused on determining which clinician would be best suited for training in ABC, based on motivation, clinical experience with families, and likelihood of remaining in Hawaii. After these conversations, the stakeholder approached 10 identified clinicians to discuss the intervention and training process, including requirements (i.e., 3-day training workshop, weekly consultation, implementation of ABC with 10 families during the year). Clinicians were given a laptop and video camera to implement the intervention, which they could keep if they completed training and consultation, which may have served as an incentive to participate. Three clinicians dropped out of training prior to beginning fidelity-focused consultation and were not included in the current study. Additional information about training and implementation in the current sample can be found in Caron et al. (2016).

All clinicians were female, and all had masters' degrees. Three clinicians were Caucasian, one was Asian American, one was Hawaiian American, and two were multi-racial. On average, clinicians were 38 years old ($SD = 8$ years, range 27–48), and had worked in their jobs for an average of 3.5 years ($SD = 2.4$, range 0.5–6.5 years) and in their fields for an average of 10.1 years ($SD = 5.8$, range 2.0–17.0). Table 1 presents details about individual clinicians'

Table 1 Descriptive data about clinicians

Clinician	# Recordings submitted during			# Consultation sessions received		# Self-coding assignments completed (% of assigned)	Average fidelity scores			
	CAU	FFC	Sustainment	CAU	FFC		On-target comments per minute	% of on-target comments	% of missed opportunities	Average components
1	17	89	54	54	40	67 (99)	1.23	80.9	56.9	0.92
2	27	61	10	41	34	56 (100)	1.03	84.7	55.3	0.85
3	26	58	7	48	26	42 (98)	1.05	87.2	62.7	1.09
4	20	156	386	40	31	74 (97)	0.93	84.2	62.2	0.86
5	20	32	n/a*	39	11	15 (88)	0.55	71.7	79.2	1.16
6	10	47	39	44	24	39 (100)	1.60	85.5	47.8	1.28
7	21	73	64	41	26	33 (92)	1.74	89.9	42.9	1.85
Average	20	74	93	44	27	47 (96)	1.16	83.4	58.1	1.14

CAU consultation as usual, *FFC* Fidelity-focused consultation

*Clinician 5 chose not to pursue certification for personal reasons, and did not continue practicing ABC after the training period

participation in study procedures, which the following section describes at the group level.

Procedure

Consultation as Usual

Clinicians completed a 3-day training workshop in March 2012. After this training, they immediately began small group clinical consultation ("consultation as usual"). Clinicians began seeing ABC cases between March and June 2012, but were expected to attend group consultation regardless of whether they had begun implementing ABC.

Consultation as usual was led by a licensed psychologist who was a former graduate student of the intervention developer. She was selected for this role because of her expertise in ABC, and had provided ABC consultation to implementation sites on a part-time basis since 2008. The consultant did not have prior experience with this group of clinicians. Consultation as usual occurred for 1 h weekly, and included session video review, feedback on implementing manual content, and discussion of case-specific dynamics. Although consultation as usual included discussion of in-the-moment commenting, the consultant was not a reliable fidelity coder. Clinicians were expected to attend group clinical consultation weekly, and on average, attended 44 group consultation sessions ($SD = 5$) during the full training period. Consultation as usual was conducted remotely using videoconferencing software.

Fidelity-Focused Consultation (with Ongoing Consultation as Usual)

Fidelity-focused consultation was intended to begin after clinicians had completed session 4 of ABC with two families. Due to differences in case load, families dropping out prior to session 4, and clinicians' speed of sending videos, they had completed 20 sessions ($SD = 6$), on average, before beginning fidelity-focused consultation. Clinicians began fidelity-focused consultation between June and September 2012, and had been implementing ABC for 3.8 months, on average ($SD = 1.2$, range 1.9–5.4 months), at that time. As in Meade et al. (2014), group clinical consultation ("consultation as usual") continued after fidelity-focused consultation was added.

Seven undergraduate students who were reliable ABC fidelity coders conducted weekly fidelity-focused consultation with clinicians. In preparation for meetings, the first author assigned clinicians coding (i.e., self-monitoring) from one to three recent session videos, with the number of assignments based on the clinician's case load and what the clinician determined was a reasonable workload. Assignments were chosen randomly by rolling a six-sided die,

with the face of the die corresponding to one of six possible segments: minutes 10–15, 15–20, 20–25, 25–30, 30–35, or 35–40. The video segment was then screened and if there were major issues that would affect the coding or representativeness of the clip (e.g., parent off-screen or on the phone), the clip was adjusted while maintaining as much of the random selection as possible. Clinicians sent their self-coding to the consultants, who sent consultant-rated coding in return. Clinicians were largely compliant with self-coding assignments; the number of missed assignments averaged 1.3 (range 0–3), representing 3.8% (range 0–12%) of the total assignments given.

Clinicians and consultants then met remotely using videoconferencing software for individual half hour sessions in which clinicians received feedback on both their self-coding accuracy and their fidelity. Feedback on self-coding was focused on refining clinicians' understanding of coding 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 included comparisons to certification criteria and prior performance, celebration of gains, and recommendations and coaching for improvement. For example, for a clinician who commented infrequently, feedback might include identification of reasons why (e.g., clinician had difficulty dividing attention between two parents) and coaching to comment more frequently in the future (e.g., "Until this gets more natural for you, just focus on noticing behaviors and commenting with one parent at a time. And if you're busy talking to mom, you could say, 'I'm sorry to interrupt, but I just have to highlight what dad is doing over there!' "). For a clinician who struggled with comment quality, feedback would likely include reviewing individual comments, explaining why the comments were coded as low quality, and discussing 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?"). Consultants also used other active learning strategies such as modeling and role-play of comments. The first author met with fidelity-focused consultants weekly for group supervision, which included review of fidelity coding and identification of areas for feedback to clinicians.

Integration of Consultation After each fidelity-focused consultation session, fidelity-focused consultants wrote brief emails to the clinical consultant about the focus of the session and their feedback. The clinical consultant would occasionally respond with information about relevant goals, feedback or case-specific dynamics discussed in consultation as usual. The clinical consultant also corresponded with the authors when clinicians asked questions about fidelity

coding or material from fidelity-focused consultation during consultation as usual.

Clinicians engaged in fidelity-focused consultation, with ongoing consultation as usual, for 10.3 months on average ($SD = 2.2$, range 7.7–14.3 months). All consultation was designed to conclude in May 2013. One clinician ended consultation early due to maternity leave in April 2013, and three clinicians ended as planned in May 2013. Three clinicians continued consultation until August or September 2013; two of these clinicians had significant breaks in implementing ABC during the fidelity-focused consultation period (3.7 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 ($SD = 41$) and 27 fidelity-focused consultation sessions ($SD = 9$), on average.

Sustainment

The sustainment period began after clinicians ended consultation. All clinicians but one were certified and continued practicing ABC after training ended. One clinician, who also had an 8.6-month break during training, did not pursue certification or continue implementing ABC for personal reasons. Clinicians were asked to send session videos for continuing program evaluation, and participation varied, with four clinicians sending most of their sessions, and two sending fewer videos. After 2 years, clinicians sent 10 recent session videos for recertification evaluation. Five clinicians submitted sessions and were recertified. Excluding one outlier clinician (386 sessions), there was an average of 35 observations ($SD = 26$) per clinician during the follow-up period. Clinicians' final follow-up data point averaged 20 months ($SD = 11$, range 2–31) after consultation ended. Excluding the outlier clinician, who had data from sessions with 70 families, the average clinician had session data nested within 14 cases ($SD = 3$, range 11–18) across the full training period.

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

Measure: ABC Fidelity

The ABC fidelity measure assesses the frequency and quality of in-the-moment commenting, a core component of ABC. The measure is coded from 5-min segments of sessions, is reliable, and predicts client outcomes (Caron et al. 2018). It involves first coding parent behaviors relevant to ABC, 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 response to the prior parent behavior. Off-target comments may be poorly timed, provide incorrect information, or highlight behaviors that are not targeted by ABC. As another measure of comment quality, the number of information components included in comments (from 0 to 3) are coded. Information components can include: (1) specifically describing the parent–child interaction ("He rolled you the ball and you rolled it back."), (2) labeling the ABC target of the parent's behavior ("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.").

Fidelity is coded on a Microsoft Excel spreadsheet that calculates summary statistics for the 5-min clip. In prior work, frequency of on-target comments and percentage of on-target comments (number of on-target comments divided by total number of comments) predicted parent behavior change at both the clinician- and case-levels (Caron et al. 2018), and were chosen as the primary outcomes. Percentage of missed opportunities (parent behaviors not commented on, divided by total opportunities to comment) and average number of components predicted behavior change only at the clinician-level and were examined as secondary outcomes. To be certified, clinicians had to demonstrate ≥ 1 on-target comment per minute or $\leq 50\%$ missed opportunities, $\geq 80\%$ on-target comments, and ≥ 1 average components, in 8 of 10 recent sessions.

Data from 1217 sessions were available and included in analyses. Excluding one outlier (562 sessions), each clinician had data from an average of 109 sessions ($SD = 42$, range 52–160). To assess reliability, 117 videos were double-coded. One-way, single measures, random effects intraclass correlations (ICCs) demonstrated excellent reliability for on-target comment frequency (0.95), percentage of on-target comments (0.76), and percentage of missed opportunities (0.76), and good reliability for average components (0.74; Cicchetti and Sparrow 1981).

For comparison, among the 311 sessions coded by both clinicians and consultants, reliability was similarly excellent for on-target comment frequency (group ICC = 0.89, individual range 0.78–0.93) and good for average components (group ICC = 0.64, individual range 0.24–0.73). However, clinicians' reliability was lower and more variable for percentage of on-target comments (group ICC = 0.55, range 0.01–0.80) and percentage of missed opportunities (group ICC = 0.26, range − 0.28 to 0.73).

Analyses

Piecewise longitudinal hierarchical linear modeling (HLM; Raudenbush and 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. Fidelity-focused consultation was split into two 5-month periods, a decision guided by Meade et al. (2014). Although Meade et al. (2014) found increasing fidelity during the 6 months after self-monitoring began, fidelity was not assessed during a longer self-monitoring period, and may or may not have continued increasing during a longer assessment period. For this reason, we split the 10-month fidelity-focused consultation period into two segments, and we hypothesized that growth would occur during the first 5 months, but we did not form specific hypotheses about the second 5 months. 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 growth during consultation as usual. Hypothesis tests compared slopes of different periods.

To provide accurate slope estimates, the 3- and 8-month breaks of two clinicians were recoded into 1-month breaks, because we did not expect to observe improvement during breaks. Additionally, one clinician who had difficulty meeting certification criteria received 4.5 more months of fidelity-focused consultation than other clinicians. To increase the reliability of group estimates, these data were recoded to fit all of the data from the clinician's final 9.5 months of fidelity-focused consultation into the second 5-month period (i.e., time coding was divided by around 2). However, changing this clinician's data did not affect 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 implementation periods. Models for each fidelity outcome (e.g., on-target comment frequency) included four slope variables, which could vary (e.g., increase) during the period of interest (e.g., consultation as usual) but which were constant during other periods. Thus, each variable estimated the expected rate of change in fidelity during consultation as usual, months 1–5 of fidelity-focused consultation, months 6–10 of fidelity-focused consultation, or sustainment.

HLM offered a flexible approach to a dataset with significant variability in the number and spacing of time points, and accounted for the nesting of time-varying session data (level 1) within clients (level 2) within clinicians (level 3; Raudenbush and Bryk 2002). Unconditional three-level models, which excluded all predictor variables, were used to estimate the variability in fidelity outcomes attributed to different levels. Most of the variance (59–93%) in the four outcomes was attributed to the time level, whereas 6–10% of the variance was attributed to the client level, and 1–31% of the variance was attributed to the clinician level. Given that clinicians were the focus of this study, clinician-level random effects terms were included for each slope and the intercept, to account for differences in clinicians' initial fidelity and trajectories of change, around the group's average intercept and estimated change. Guided by Snijders' (2005) recommendations to keep the model parsimonious and avoid overfitting, we included one client-level error term 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 depicted in Fig. 1, over the course of the first 5 months of fidelity-focused consultation, on average, clinicians showed an estimated increase of 0.76 comments per minute and 17.8% on-target comments. As shown in Table 2, 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 of the primary outcomes. As shown in Table 2 and Fig. 2, during the first 5 months of fidelity-focused consultation, clinicians showed decreasing missed opportunities to comment, change that was not observed during other periods. This result parallels that found for comment frequency, showing that clinicians were commenting more frequently and less frequently failing to respond to relevant parent behaviors after fidelity-focused consultation began.

For the average number of components in comments, no slopes were significantly different from 0. However, the rate of change during consultation as usual approached significance ($p = .066$) and was significantly steeper than the slope of the sustainment period. Overall, clinicians appeared to make fairly steady gains across all three training periods, before plateauing in the sustainment period. When data were recoded into a two-piece model (training vs. sustainment), the training period slope, $\beta_{10} = 0.04$, still only approached significance ($p = .055$).

Fig. 1 Primary outcomes. *p < .05, **p < .01. Consultation as usual occurred from months −5 to 10; fidelity-focused consultation was added from months 0 to 10. Graphs show the first year of the sustainment period, but data were modeled through 31 months of sustainment
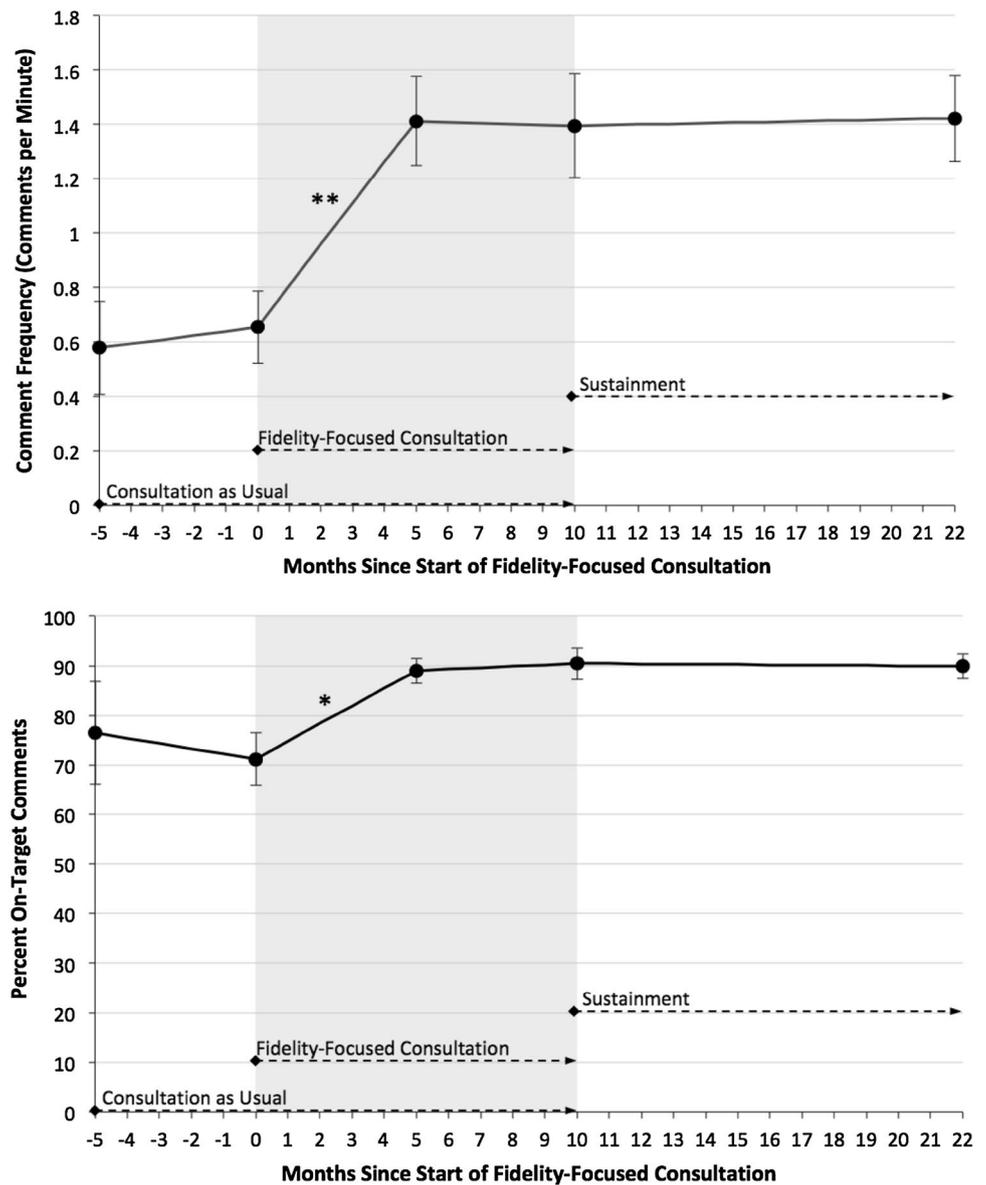
Thus, for three of four fidelity outcomes, the periods of largest growth occurred during the first 5 months of fidelity-focused consultation. In contrast, for these three outcomes, during the consultation as usual training period, minimal change was observed. There was also minimal change during the 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 for up to 30 months without additional consultation.

For exploratory purposes, data were also modeled in three-piece models, in which the fidelity-focused consultation period was represented by only one slope. The fidelity-focused consultation slopes for on-target comment frequency, percentage of on-target comments, and percentage of missed opportunities remained significantly different from 0, but were less steep than the slopes for the first 5 months of fidelity-focused consultation found in the four-piece models ($\beta_{20} = 0.08$ for on-target comment frequency, $\beta_{20} = 1.98$ for percentage of on-target comments, and $\beta_{20} = -1.90$ for percentage of missed opportunities). As in the four-piece models, the slopes for the consultation as usual and sustainment periods were not significantly different from 0.

Discussion

The current study examined the impact of fidelity-focused consultation, in which clinicians engaged in self-monitoring and received fidelity feedback from an expert coder on a weekly basis. The results suggest that this consultation

Table 2 Hierarchical linear models testing growth in fidelity during different periods of implementation

Effects	Coefficients	SEs	t-ratios	p-values
Comment frequency (number of on-target comments per minute)				
Intercept, β_{00}	0.65	0.13	4.97	0.003
Consultation as usual, β_{10}	0.02	0.05	0.33	0.75
Fidelity consultation months 1–5, β_{20}	0.15	0.03	5.08	0.002
Fidelity consultation months 6–10, β_{30}	− 0.00	0.03	− 0.12	0.91
Sustainment, β_{40}	0.00	0.01	0.32	0.76
Hypothesis tests: $\beta_{20} > \beta_{10}$ ($p < .05$), $\beta_{20} > \beta_{30}$ ($p < .01$), $\beta_{20} > \beta_{40}$ ($p < .001$)				
Percentage of on-target comments				
Intercept, β_{00}	71.19	5.29	13.45	< 0.001
Consultation as usual, β_{10}	− 1.06	2.75	− 0.38	0.71
Fidelity consultation months 1–5, β_{20}	3.56	1.05	3.39	0.015
Fidelity consultation months 6–10, β_{30}	0.29	0.98	0.30	0.78
Sustainment, β_{40}	− 0.05	0.15	− 0.31	0.77
Hypothesis tests: $\beta_{20} > \beta_{30}$ ($p < .05$), $\beta_{20} > \beta_{40}$ ($p < .01$)				
Average number of comment components				
Intercept, β_{00}	1.05	0.10	10.44	< 0.001
Consultation as usual, β_{10}	0.06	0.03	2.25	0.066
Fidelity consultation months 1–5, β_{20}	0.04	0.03	1.23	0.27
Fidelity consultation months 6–10, β_{30}	0.06	0.03	1.73	0.13
Sustainment, β_{40}	− 0.01	0.00	− 1.98	0.10
Hypothesis tests: $\beta_{10} > \beta_{40}$ ($p < .05$)				
Percentage of missed opportunities				
Intercept, β_{00}	68.24	3.74	18.26	< 0.001
Consultation as usual, β_{10}	− 0.74	1.26	− 0.59	0.58
Fidelity consultation months 1–5, β_{20}	− 2.83	0.77	− 3.67	0.011
Fidelity consultation months 6–10, β_{30}	− 0.91	1.05	− 0.86	0.42
Sustainment, β_{40}	0.07	0.17	0.39	0.71
Hypothesis tests: $\beta_{20} > \beta_{40}$ ($p < .001$)				

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

procedure may have increased clinicians' growth in ABC fidelity over time. Although the results should be viewed as preliminary, given the study design, they are consistent with other studies that have found that performance feedback using fidelity data can result in continued improvement following training (Eiraldi et al. 2018; Weck et al. 2017). In addition, these results were found when fidelity-focused consultation was added to weekly hour-long "consultation as usual" using video review of recent sessions, representing a "consultation as usual" protocol more intensive than in some prior work (e.g., Eiraldi et al. 2018). Thus, although the current study's design and sample size limit the conclusions that can be drawn, the results offer an exciting signal about what might be found in future work.

One unique aspect of this study's feedback intervention, compared with most prior work, is that clinicians coded their own fidelity. Additionally, for some of the fidelity variables, clinicians had good interrater reliability with consultants. Accurate self-coding may have increased clinicians' receptivity to consultants' feedback by increasing perceived source credibility and feedback validity, which may have enhanced the feedback's impact (Riemer et al. 2005). In addition, given that some consultants may avoid giving objective, coding-based feedback because they believe it may be perceived as harsh by clinicians (Beckman et al. 2017), having clinicians complete self-coding may make it easier for consultants to provide such feedback. Further, the process of learning to code fidelity may build clinicians' knowledge of treatment fidelity and ability to process client behavior that transfer to intervention sessions (Isenhart et al. 2014). Thus, self-monitoring likely contributed to the apparent impact of fidelity-focused consultation.

To some, the finding that fidelity coding feedback leads to changes in fidelity may seem like "teaching to the test." In fact, after Dunn et al. (2016) found that their less structured, lower cost feedback intervention did not increase MI 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 is what is needed to raise [fidelity] scores over time" (p. 81).

Fig. 2 Secondary outcomes. *p < .05. Consultation as usual occurred from months −5 to 10; fidelity-focused consultation was added from months 0 to 10. Graphs show the first year of the sustainment period, but data were modeled through 31 months of sustainment
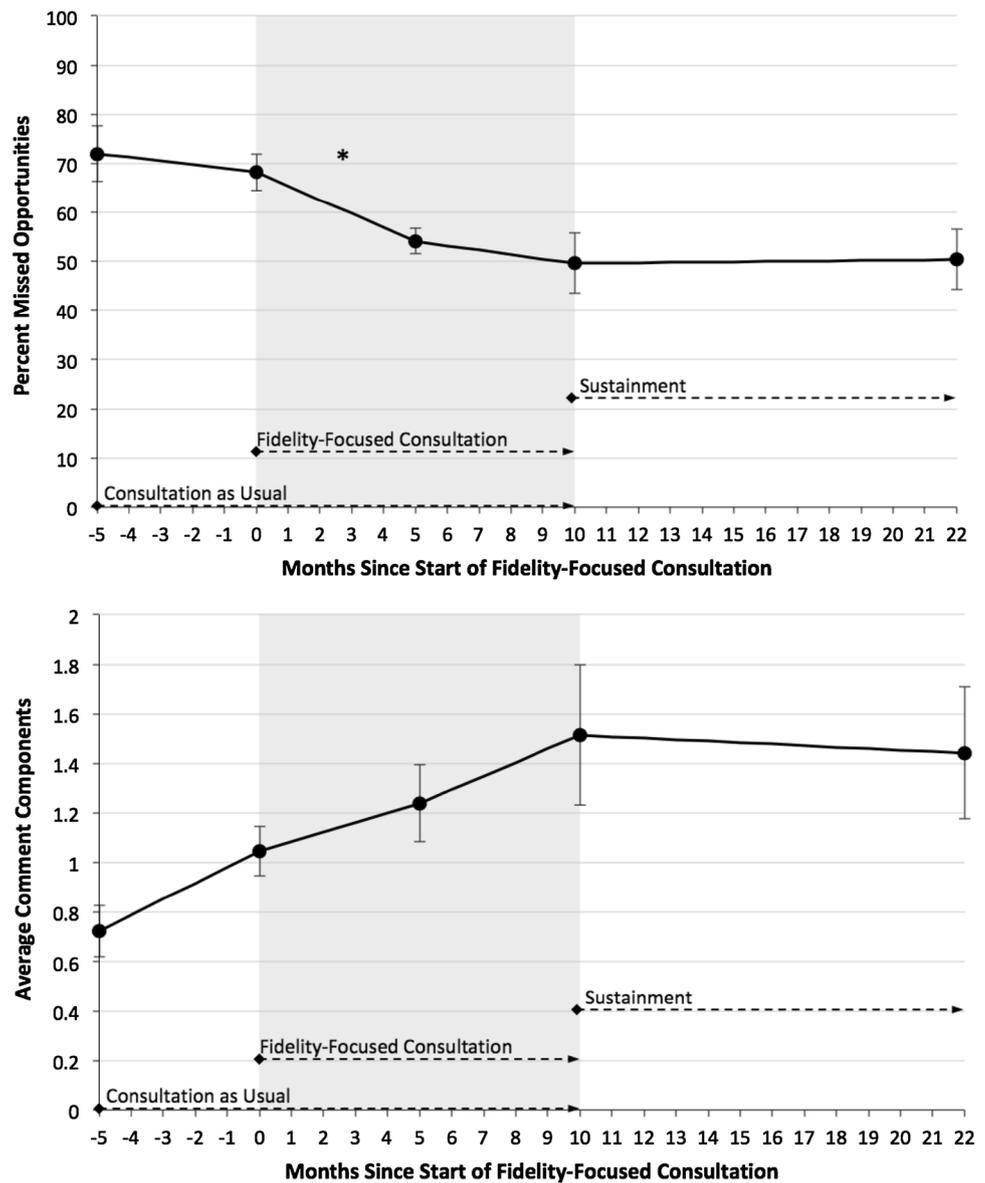
Although "teaching to the test" has a negative connotation, this undertone does not apply to fidelity feedback interventions, and to these results in particular, for two reasons. First, the "test" has been linked to desired outcomes for families (Caron et al. 2018), showing its value. Second, clinicians maintained performance on the "test" for over 2 years after they stopped receiving "grades," showing that their learning 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 most significant strengths. Skill levels following training and consultation have been maintained for 6 to 15 months in studies of clinicians learning Multisystemic Therapy (Henggeler et al. 2008), measurement-based care (Lyon et al. 2017),

MI (Persson et al. 2016), and Trauma-Focused Cognitive Behavioral Therapy (Murray 2017). The current study's 2-year sustainment period with its demonstrated lack of deterioration in fidelity is rare in the literature, and supports the effectiveness of ABC consultation procedures. Although clinicians were encouraged to continue self-coding fidelity after consultation ended, it is unknown whether they actually did so. Continuing to self-monitor using fidelity coding may have maintained fidelity during the sustainment period, but clinicians also could have internalized aspects of the coding process, which could have supported fidelity through "live" self-monitoring.

The current study also found a lack of growth in fidelity during the second 5 months of fidelity-focused consultation,

sparking questions about the necessary dose of such consultation. Although the current results suggest that clinicians may not need a full 10 months to achieve the levels of ABC fidelity required for certification, it is possible that the second 5 months of fidelity-focused consultation are necessary to achieve the observed long-term sustainment of fidelity. 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 represent the full operation stage, when new learning is integrated into clinicians' practice and implemented competently. It may be important for clinicians to experience the full operation stage while still receiving expert consultation in order to consolidate previously learned skills and refine self-monitoring practices they can use after consultation ends. 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 intervene with fewer clients may need more time to reach certification requirements. Indeed, the results presented here are group patterns; as described in the "Method" section, one clinician required additional consultation to meet certification requirements.

The current study also raised questions about the importance and necessary dose of consultation as usual. Because consultation as usual was a constant throughout the training period, it is possible that fidelity-focused consultation only leads to increased fidelity when clinicians are receiving traditional consultation concurrently. Additionally, given the study's multiple-baseline design, it is also possible that order/timing effects are present, and the response to fidelity-focused consultation is only observed after clinicians have received traditional consultation for several months.

Questions about necessary doses of consultation are important because the more cost effective that ABC training can be, the more easily and widely it can be implemented (Weisz et al. 2011). Two aspects of fidelity-focused consultation lower its cost. First, only 5-min video clips are coded; typically, this coding can be completed in about 30 min. Second, all of the fidelity-focused consultants in the current study were undergraduate coders who received course credit as compensation for their time, showing that advanced degrees are not required for effective support. 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 clinical Ph.D.-level consultants (Caron et al. 2016).

A limitation of the current results is that only one of four outcomes, comment frequency, showed statistically stronger growth during fidelity-focused consultation than during 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 3 and a significant amount of unexplained, between-clinician variance in the consultation as usual slope. Additionally, the consultation as usual period was shorter and had fewer observations per clinician than other periods, which also may have lowered power to detect differences between this slope and those of other periods.

A major weakness of the current study is the small sample size at level 3, which limited statistical power and generalizability. Generalizability is also limited because only 7 of the original 10 trainees participated in fidelity-focused consultation, and only 6 participated in the sustainment phase of the study, so the results only reflect what might be expected of ABC trainees who complete training and are certified. A further limitation to generalizability is the fact that the sample was selected for its motivation and experience with families, and the clinicians were highly compliant with self-coding. Future work in larger samples should examine the effects of engagement and compliance on fidelity-focused consultation.

In the current sample, randomization to consultation conditions would have further reduced power by splitting the sample into smaller groups, but because there was no control group, 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. However, the variability in the onset of fidelity-focused consultation makes this possibility somewhat less likely. Prior studies of consultation procedures combining fidelity feedback and self-monitoring of fidelity have also had non-experimental designs and small sample sizes ($n = 2$ in Isenhart et al. 2014, and $n = 20$ with effects observed only in a selected group of $n = 4$ in Garcia et al. 2017). In the current study, the number of observations per clinician was a strength, and allowed for nuanced modeling of fidelity across time, and reliable estimates of clinicians' behavior.

This study was designed to test the impact of fidelity coding feedback and self-monitoring as a consultation intervention. Although the intervention as a package appeared to promote fidelity, future research should test it in a larger and more rigorous, randomized controlled trial with conditions including consultation as usual, fidelity-focused consultation, and their combination. Designing and testing similar fidelity-focused consultation interventions for other

evidence-based treatments would also increase confidence in the generalizability of the findings. Of note, the ABC fidelity measure and the MI fidelity measure used by Isenhart et al. (2014) focus clinicians' attention on very specific verbal behaviors; it would be interesting to test whether fidelity-focused consultation using a fidelity measure with broader, global scales could also be effective.

Future research should also explore the active ingredients of fidelity-focused consultation. For example, the impact of self-monitoring could be separated out and tested in a trial in which clinicians are randomized to fidelity-focused consultation with and without self-monitoring. Additionally, consultation sessions could be videotaped and examined for provision of fidelity feedback and other potential active ingredients of consultation (e.g., modeling, role-play), which could be tested as predictors of change in clinicians' fidelity. Further, because clinicians were able to achieve acceptable reliability on some of the fidelity outcomes, future research should explore the potential utility of clinicians' self-ratings. Clinicians' self-coding accuracy should be explored as an outcome of consultation, as well as a potential predictor of fidelity growth and indicator of necessary dosage of consultation.

In summary, this study joins a fairly small literature showing that consultation procedures that incorporate fidelity coding feedback promote implementation outcomes (Eiraldi et al. 2018; Miller et al. 2004; Weck et al. 2017). It joins a smaller literature that suggests added benefit for self-monitoring of fidelity (Garcia et al. 2017; Isenhart et al. 2014). To contextualize these results in 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 study. This study established fidelity-focused consultation as a communication link process and potential active ingredient of implementation through its apparent impact on clinicians' fidelity, over and above traditional consultation. A unique aspect of Fixsen et al.'s (2005) model 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-monitoring of fidelity, can provide a feedback loop to clinicians themselves, which may serve as an active ingredient of implementation. One strength of the study is that the impact of fidelity-focused consultation was examined across multiple stages defined by Fixsen et al. (2005), from initial implementation through sustainability.

Compliance with Ethical Standards

Conflict of interest EB Caron declares that she has no conflict of interest. Mary Dozier declares that she has no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

References

Beckman, M., Bohman, B., Forsberg, L., Rasmussen, F., & Ghaderi, A. (2017). Supervision in motivational interviewing: An exploratory study. *Behavioural and Cognitive Psychotherapy, 45*, 351–365. https://doi.org/10.1017/s135246581700011x.

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. https://doi.org/10.1176/appi.ps.201100401.

Bernard, K., Dozier, M., Bick, J., Lewis-Morrarty, E., Lindheim, O., & Carlson, E. (2012). Enhancing attachment organization among maltreated children: Results of a randomized clinical trial. *Child Development, 83*, 623–636. https://doi.org/10.1111/j.1467-8624.2011.01712.x.

Bernard, K., Hostinar, C. E., & Dozier, M. (2015a). 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. https://doi.org/10.1001/jamapediatrics.2014.2369.

Bernard, K., Simons, R., & Dozier, M. (2015b). Effects of an attachment-based intervention on child protective services-referred mothers' event-related potentials to children's emotions. *Child Development, 86*, 1673–1684. https://doi.org/10.1111/cdev.12418.

Caron, E., Bernard, K., & Dozier, M. (2018). In vivo feedback predicts parent behavior change in the Attachment and Biobehavioral Catch-up intervention. *Journal of Clinical Child and Adolescent Psychology, 47*, S35–S46. https://doi.org/10.1080/15374416.2016.1141359.

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

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, 86*, 127–137.

Darnell, D., Dunn, C., Atkins, D., Ingraham, L., & Zatzick, D. (2016). A randomized evaluation of motivational interviewing training for mandated implementation of alcohol screening and brief intervention in trauma centers. *Journal of Substance Abuse Treatment, 60*, 36–44. https://doi.org/10.1016/j.jsat.2015.05.010.

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

Dunn, C., Darnell, D., Atkins, D. C., Hallgren, K. A., Imel, Z. E., Bumgardner, K., et al. (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. (2013). Dissemination and implementation of evidence-based practices: Training and consultation as implementation strategies. *Clinical Psychology: Science and Practice, 20*, 152–165. https://doi.org/10.1111/cpsp.12031.

Eiraldi, R., Mautone, J. A., Khanna, M. S., Power, T. J., Orapallo, A., Cacia, J., et al. (2018). Group CBT for externalizing disorders in urban schools: Effect of training strategy on treatment fidelity and child outcomes. *Behavior Therapy, 49*, 538–550. https://doi.org/10.1016/j.beth.2018.01.001.

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, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network. https://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature.

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 and Adolescent Psychology, 44*, 730–741. https://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, 29*, 313–325. https://doi.org/10.1080/10401334.2017.1279056.

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

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. https://doi.org/10.1080/19345740802539325.

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. https://doi.org/10.3109/14659891.2013.765514.

Jamtvedt, G., Young, J. M., Kristoffersen, D. T., O'Brien, M. A., & Oxman, A. D. (2006). Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *BMJ Quality and Safety, 15*, 433–436. https://doi.org/10.1136/qshc.2006.018549.

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.

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 and Adolescent Psychology*. https://doi.org/10.1080/15374416.2017.1280808.

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 and Human Development, 16*, 356–370. https://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. https://doi.org/10.1037/0022-006X.72.6.1050.

Murray, H. (2017). Evaluation of a trauma-focused CBT training programme for IAPT services. *Behavioural and Cognitive Psychotherapy, 45*, 467–482. 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. https://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, MD: National Institutes of Health, National Institute of Mental Health. https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/blueprint-for-change-research-on-child-and-adolescent-mental-health.shtml.

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.

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

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. https://doi.org/10.1016/j.chc.2004.05.002.

Roben, C. K., Dozier, M., Caron, E., & Bernard, K. (2017). Moving an evidence-based parenting program into the community. *Child Development, 88*, 1447–1452. https://doi.org/10.1111/cdev.12898.

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. https://doi.org/10.1207/S15374424JCCP3301_10.

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

Snijders, T. A. B. (2005). Fixed and random effects. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 664–665). Chichester: Wiley.

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. https://doi.org/10.1037/a0018544.

Webster-Stratton, C. H., Reid, M. J., & Marsenich, L. (2014). Improving therapist fidelity during implementation of evidence-based practices: Incredible Years Program. *Psychiatric Services, 65*, 789–795. https://doi.org/10.1176/appi.ps.201200177.

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. https://doi.org/10.1080/10503307.2015.1132857.

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. https://doi.org/10.1111/j.1468-2850.2010.01233.x.

Yarger, H. A., Bernard, K., Caron, E., Wallin, A., & Dozier, M. (2019). Enhancing parenting quality for young children adopted internationally: Results of a randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology.* https://doi.org/10.1080/15374416.2018.1547972.