Comparison of Video and Audio Rating Modalities for Assessment of Provider Fidelity to a Family-Centered, Evidence-Based Program

By

Jenna Rudo-Stern

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2018 by the
Graduate Supervisory Committee:

Sharlene Wolchik, Chair
Anne Mauricio
Kevin Grimm
Nancy Gonzalez

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

The current study assessed whether the interrater reliability and predictive validity of fidelity ratings differed significantly across the modalities of audio and video recordings. As empirically supported programs are moving to scale, attention to fidelity, the extent to which a program is delivered as intended, is essential because high fidelity is needed for positive program effects. Consequently, an important issue for prevention science is the development of feasible and acceptable methods for assessing fidelity. Currently, fidelity monitoring is rarely practiced, as the typical way of measuring fidelity, which uses video of sessions, is expensive, time-consuming, and intrusive. Audio recording has multiple advantages over video recording: 1) it is less intrusive; 2) equipment is less expensive; 3) recording procedures are simpler; 4) files are smaller so it takes less time to upload data and storage is less expensive; 5) recordings contain less identifying information; and 6) both clients and providers may be more willing to have sensitive interactions recorded with audio only. For these reasons, the use of audio recording may facilitate the monitoring of fidelity and increase the acceptability of both the intervention and implementation models, which may serve to broaden the scope of the families reached and improve the quality of the services provided. The current study compared the reliability and validity of fidelity ratings across audio and video rating modalities using 77 feedback sessions drawn from a larger randomized controlled trial of the Family Check-Up (FCU). Coders rated fidelity and caregiver in-session engagement at the age 2 feedback session. The composite fidelity and caregiver engagement scores were tested using path analysis to examine whether they predicted parenting behavior at age 3. Twenty percent of the sessions were double coded to assess interrater reliability.

The interrater reliability and predictive validity of fidelity scores and caregiver engagement did not significantly differ across rating modality. However, caution must be used in interpreting these results because the interrater reliabilities in both conditions were low. Possible explanations for the low reliability, limitations of the current study, and directions for future research are discussed.

This dissertation is dedicated to my husband, our family, my cohort, and the Swim Team. Thank you for traveling this path with me and for offering so much love, music, physical activity, consultation, and humor along the way.

This dissertation is also in memory of my graduate advisor, Dr. Thomas J. Dishion.

ACKNOWLEDGMENTS

I wish to express my appreciation for the generosity shown to me by my graduate advisor, Dr. Thomas Dishion, who greatly facilitated my entry into the fields of prevention and implementation science, for which we shared a passion. I offer my immense gratitude to Dr. Sharlene Wolchik, who chaired the dissertation committee and provided extensive feedback throughout the writing process. Her integrity as a scientist is a model for me. Dr. Anne Mauricio has served as my primary mentor and collaborator for the past five years, and I hope this is merely the beginning of a lifelong association. I am also grateful to Drs. Nancy Gonzalez and Kevin Grimm for their guidance and contributions to this dissertation project. Dr. Justin D. Smith was generous with his knowledge of previous studies on the COACH, and I offer him my sincere thanks. I wish also to acknowledge the contributions of two staff at the Oregon Research Institute: Jenene Peterson, who spent hours organizing and delivering the intervention session recordings used for this study, and Shannon McGill, who gathered and shared assessment reports and numerous data files and who tirelessly sleuthed out answers to my various questions. Of course, I owe enormous thanks to the coding team: Cameron Bond, Kiana Frutiger, Mitch Gibson, Laura Hemphill, and Connor Seiferth. Thanks also to Amelia Stahl for conducting the orchestra of coding assignments. This study would not have been possible without the endless hours of work contributed by these six excellent team mates. I would like also to acknowledge that data used in this study was collected for research supported by National Institute on Drug Abuse grant DA016110, awarded to Dr. Thomas Dishion, Daniel Shaw, Melvin Wilson, and Frances Gardner.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Introduction

Fidelity, the extent to which the program is delivered as intended (Carroll et al., 2007), is essential to the successful transfer of empirically supported programs to implementation in community service delivery systems (Breitenstein, Gross, et al., 2010). Fidelity assessment during the implementation process also can be a tool for providing feedback to improve providers' performance and ultimately increase intervention effects on targeted outcomes. However, fidelity monitoring is rarely practiced (Durlak & DuPre, 2008; Dusenbury, Brannigan, Falco, & Hansen, 2003; Dusenbury, Brannigan, Falco, & Lake, 2004), and fidelity monitoring procedures are rarely reported (Breitenstein, Gross, et al., 2010; Mihalic, 2004). Further, the typical way of measuring fidelity, which involves video recording of sessions, is expensive and time-consuming (Perepletchikova, Treat, & Kazdin, 2007), and few reliable and valid measures of fidelity, particularly for large-scale dissemination projects, exist (Breitenstein, Gross, et al., 2010). The low level of monitoring fidelity may be in part due to a dearth of pragmatic fidelity-monitoring practices and measures.

While costly, observational coding systems implemented by trained coders remain the gold standard for assessing fidelity (Gearing et al., 2011; Gillespie, Huey, & Cunningham, 2017). One avenue for curbing associated costs and increasing feasibility and acceptability of direct observation is to use audio recordings instead of video recordings, as video recordings are more expensive and pose a greater potential threat to privacy. Focus group data indicated that providers who were asked to video record their sessions for training purposes experienced multiple frustrations, including difficulties

uploading the recordings to a secure server to be reviewed for fidelity (Mauricio, Rudo-Stern, Chiapa, et al., under review; Mauricio, Rudo-Stern, Dishion, Letham, & Lopez, under review; Mauricio, Rudo-Stern, Dishion, Shaw, et al., under review). It was common for providers to describe the uploading process as prohibitively slow, with associated costs. In addition to technical difficulties with video recording and uploading, acceptability of being video recorded may be low for both providers and their clients. Supervisors in these focus groups reported that providers are "not used to being videotaped" and that the review of these tapes led to "trepidation" and was perceived as "difficult," impeding uptake (i.e., providers do not use the model because, for training purposes, they are expected to video record their sessions). Providers also voiced concerns about client discomfort, with one participant saying, "People don't want the videotaping." These concerns about their own and client discomfort as well as technical difficulties are consistent with findings from qualitative research with medical trainees in regard to video recording their patient encounters (Eeckhout, Gerits, Bouquillon, & Schoenmakers, 2016).

Audio recording has multiple advantages over video recording: 1) it is less intrusive; 2) equipment is less expensive; 3) procedures are simpler and training is less intensive; 4) files are smaller so it takes less time to upload data and storage is less expensive; 5) recordings contain less identifying information; and 6) both clients and providers may be more comfortable and more willing to have sensitive interactions recorded with audio-only than with video. For these reasons, the use of audio recording instead of video recording has the potential to facilitate the monitoring of fidelity and

increase the acceptability of both the intervention and implementation models that require monitoring of fidelity, which may serve to broaden the scope of the families served and improve the quality of the services provided.

Video and audio recording have both been used for rating fidelity with high interrater reliability. Forgatch and colleagues (2005) evaluated fidelity to the Parent Management Training – Oregon Model and Eames and colleagues (2008) evaluated fidelity to the Incredible Years parenting program by reviewing videotapes of sessions and found high interrater reliability, with intraclass correlations ranging from 0.71 to 0.99. Dumas and colleagues (2001) reviewed audio recordings to assess process and content fidelity (i.e., adherence and competence) to the EARLY ALLIANCE prevention program family sessions and obtained interrater agreement of 87% on the measure of adherence and 97% on the measure of competence. Breitenstein and colleagues (2010) reviewed audio recordings to assess fidelity to the Chicago Parenting Program. They measured adherence and competence and obtained high interrater agreement (Adherence Scale = 94%, Competence Scale = 85%) and good to excellent intraclass correlation coefficients (Adherence Scale = 0.69, Competence Scale = 0.91). Gillespie and colleagues (2017) reviewed audio recordings to assess fidelity to Multisystemic Therapy and found that reliability of fidelity ratings ranged from fair (ICC = 0.55) to excellent (ICC = 0.76) within coder dyads and was good when calculated across all coders (ICC = 0.64). Both video- and audio-based ratings of fidelity have been shown to predict intervention outcomes (Forgatch et al., 2005; Gillespie et al., 2017).

Literature comparing the audio and video modalities for rating provider interactions with clients is sparse and has been conducted primarily in medical settings (Dent, Brown, Dowsett, Tattersall, & Butow, 2005; Riddle et al., 2002; Weingarten, Yaphe, Blumenthal, Oren, & Margalit, 2001; Williams, Herman, & Bontempo, 2013). Williams and colleagues (2013) compared audio and video rating modalities for a measure of nursing staff communication with patients. They used two different groups of coders, each group assigned to one of the rating modalities. Twenty coders reviewed 20 unique, 1-minute recordings of bathing care interactions as video and rated staff communication on 12 descriptors related to how the staff conveyed care, respect, and control. A second group of 20 coders reviewed the same clips as audio recordings. Interrater reliability within each group was excellent: ICC was 0.91 for audio and 0.94 for video. Within-person item correlations agreed across modalities and exploratory factor analysis showed comparable solutions with similar loadings for the two modalities. They concluded that visual cues such as gestures and facial expressions did not significantly change scores.

Weingarten and colleagues (2001) used one coder to compare audio and video rating modalities for a patient-centeredness scale. The coder rated the first 2 minutes of 258 patient visits, across 47 doctors, first by listening with the image off. The same coder then rated these clips based on audio-video review. For each review, the coder identified patient "offers" and then classified the doctor's response on a 4-point scale: ignored (0), closed answer or prevent further discussion (1), encouragement (2), or facilitation (3). The patient-centeredness score was computed by totaling the points then dividing by the

number of patient "offers." The mean scores were similar across modalities: video mean score was 1.94, *SD* = 0.59; audio mean 1.94, *SD* = 0.63. The differences between audio and video scores were plotted by the mean of the 2 scores. The researchers found no clinically significant difference between the rating modalities and concluded that audio review of these doctor-patient interactions was equivalent to video review.

Dent and colleagues (2005) also used only one coder to compare audio and video scores on measures of content (i.e., discussion of diagnosis, treatment, psychosocial issues, etc.) and function (e.g. inform/educate, label/judge/criticize, check patient understanding, etc.). They recorded one-on-one interactions between 10 oncologists and an actor across three different patient engagement conditions. The coder first rated the 30 recordings as audio. Three months later, the same coder rated 10 of those recordings, randomly selected, as video. The weighted Kappa scores within the same coder for comparison of audio and video coding were 0.77 for content and 0.72 for function. The authors concluded that audio recording is adequate for evaluating the doctor-patient interaction.

Riddle and colleagues (2002) rated recordings of interactions between 47 patients and 12 oncologists on interaction-level variables such as hierarchical rapport (the extent to which the doctor exhibits arrogance or cordiality to assert her or his status as a medical expert), doctor responsiveness to patient concerns; doctor dominance of floor time (the degree to which the doctor talks vs. engaging in conversational turn-taking), and connectedness/closeness (the extent to which there appears to be a warm relationship between the doctor and patient during the interaction). The study compared two rating

systems, the MAAS and the RIAS, as well as rating modalities. Each doctor-patient interaction was rated eight times by different coders (2 measures x 2 modalities; two coders per condition). Reliability estimates for both rating modalities were high (mean ICCs ranging from 0.78 to 0.89), and the reliability did not differ significantly across modalities. There were differences in the underlying factor structure of the two rating systems when compared across modalities. Additionally, many of the scores for audio ratings were higher. However, the authors concede that "if the research task is to examine some type of verbal behavior or conversational structure (e.g., question-asking, forms of address, specific information-giving) and the tone of that behavior, audiotaping is likely to be sufficient" (p.236).

The limited research comparing audio and video rating modalities for measuring provider-patient interactions in a medical setting found no significant differences in interrater or intrarater reliability. However, the methodology used in the studies raises questions about the validity of the findings. For example, using two separate groups of coders to code using one rating modality introduces coder differences as a potential confound, and using the same coder to rate both audio and video recordings introduces carryover effects. Further, these studies did not measure treatment fidelity per se, nor did they examine the predictive validity of the codes derived from audio versus video recordings.

This study addresses this gap in the literature by comparing both the interrater reliability and predictive validity of a fidelity rating system for a parenting intervention across audio and video rating modalities. This issue is of particular importance given that

many parenting interventions that have shown positive effects in efficacy trials are going to scale and require acceptable and feasible systems for monitoring fidelity to increase the likelihood that they will have the same positive impact as demonstrated in the original studies (Committee on the Prevention of Mental Disorders and Substance Abuse Among Children Youth and Young Adults: Research Advances and Promising Interventions, 2009). This study addresses this research question using data from the Family Check-Up (FCU), an evidence-based program currently in the process of scale-up (Mauricio, Rudo-Stern, Dishion, Letham, et al., under review).

Fidelity to the FCU protocol is assessed with the COACH, an observational system developed to evaluate provider adherence to the FCU and the competence with which the model is delivered. As measured by the COACH, adherence refers to the degree to which providers implementing the intervention follow the intervention protocol, delivering the content according to the structure and approach prescribed by the FCU intervention model and avoiding what is proscribed. Competence is defined as the skill with which the intervention is delivered and includes the use of clinical and process skills to promote engagement and behavior change. In addition to measures of these aspects of fidelity, the COACH assesses caregiver engagement. Caregiver engagement is viewed as an important dimension in conceptual models of implementation (Berkel, Mauricio, Schoenfelder, & Sandler, 2011). Caregivers' active participation in sessions has been associated with program effects across multiple intervention models, and competent delivery of interventions has been associated with client/caregiver participation engagement (Berkel et al., 2011; Haine-Schlagel & Walsh, 2015;

Perepletchikova & Kazdin, 2005). To date, review of video recordings of sessions has been used for rating both fidelity to the FCU and caregiver engagement.

The COACH has demonstrated adequate interrater reliability and predictive validity (Dishion, Smith, Gill, Shaw, & Knutson, 2014). The COACH, when used to rate sessions with families of 2-year-olds with high levels of problem behavior, had intraclass correlation coefficients (ICC) of 0.57-0.76 on the individual domains and an ICC for the composite score of 0.74 (Smith, Dishion, Shaw, & Wilson, 2013). Another study reported acceptable ICCs for the composite scores with this same sample when the children were age 3 (0.73), age 4 (0.77), and age 5 (0.71) (Chiapa et al., 2015). Additionally, ratings of caregiver engagement has good interrater reliability (ICC range: 0.80-0.87) (Dishion et al., 2014).

Smith and colleagues (2013) found that greater fidelity to the FCU in sessions when children were 2 years old was related to improvement in parent-reported problem behaviors at age 4 and that this effect was mediated by ratings of in-session caregiver engagement at the age 2 feedback session and by improvements in positive behavior support by parents at the age 3 assessment. Chiapa and colleagues (2015) examined drift in fidelity to the FCU over a four-year period and found that declines in fidelity were associated with less improvement in child problem behaviors.

Although Smith and his colleagues (2013) found overall reliability of the fidelity scores to be acceptable, they identified some low reliability estimates (ICC = 0.59 for Conceptual accuracy and ICC = 0.57 for Careful when teaching). Competence has been identified as a particularly difficult component of fidelity to measure, with most

instruments demonstrating low interrater reliability for competence scales relative to scales measuring adherence (Breitenstein, Gross, et al., 2010). Because a key component of the FCU is that case conceptualization and treatment recommendations be based on the FCU child and family assessment, providers are expected to act in accordance with family assessment results in order to deliver the model with competence. Previous research has shown that coders attained higher interrater reliability on the COACH when provided with family assessment data prior to rating a session (Smith et al., 2016). Thus, the FCU developer revised the COACH to have coders review the family's assessment data prior to rating the session. He also added non-exemplars of fidelity to both the rating form and manual to provide greater clarity around proscribed provider behaviors.

Current Study

As evidence-based parenting programs are being increasingly disseminated (Sanders & Kirby, 2012), it is important to ensure that these programs are delivered with high fidelity to increase the likelihood that they have the intended effects (Breitenstein, Gross, et al., 2010). The availability of methods of assessing fidelity that are not only valid and reliable but also feasible and acceptable is required for ongoing fidelity monitoring to prevent drift. Given that providers and agencies are reluctant to use video recording, either because of concerns regarding privacy or because of the associated expense, the use of audio recording as an alternative could increase adoption of parenting programs that require ongoing fidelity monitoring for successful implementation. The use of audio review may increase the acceptability of these programs' implementation

models and the likelihood that fidelity is monitored and ratings are used to improve implementation and, ultimately, client outcomes.

Based on previous studies comparing the interrater reliabilities of the audio and video modalities for rating provider-patient interactions in medical settings, I hypothesized that the interrater reliability of the COACH would not be significantly different across rating modalities. Additionally, given that the fidelity criteria focus on verbal behavior, conversational structure, and tone, I hypothesized that review of audio recording was likely to yield composite fidelity scores not significantly different from those based on video review and that the predictive validity of the COACH would not be significantly different across the two rating modalities. This comparison was made using the following model: treatment fidelity will predict caregiver engagement at the age 2 feedback session and caregiver engagement at the age 2 feedback session will predict observed positive behavior support at the age 3 assessment.

Methods

Participants

This study used data from a subsample of families in a randomized trial of FCU (Dishion et al., 2008). Families were recruited from the Women, Infants, and Children Nutritional Supplement Program in three culturally and geographically diverse regions in the U.S.: Eugene, OR (271 dyads); Charlottesville, VA (188); and Pittsburgh, PA (272). Caregivers with a 2-year-old child were screened for risk factors for future child behavior problems and those families that were deemed at-risk were invited to enroll in the study. Participants were randomly assigned to either a services-as-usual condition or to a group

that was offered the Family Check-Up annually. The providers delivering the Family Check-Up had either a master's or doctoral degree, were trained in the Family Check-Up, participated in group supervision, and received ongoing fidelity monitoring and feedback throughout the course of the study.

Families in the current study were a subsample of the 731 caregivers (96% mothers) who participated in the trial. This subsample was nearly identical to the sample used in the Smith et al. (2013) validation study of an earlier version of the COACH and was selected in order to use the caregiver engagement data collected by the previous coding team as part of the current analyses. For two families in that sample, only audio was available, so those two families were dropped from the current analyses, although their sessions were coded.[1] The 77 families (OR, 28; VA, 23; PA, 26) were assigned to the FCU condition, received the FCU feedback session at age 2, had a recording of this session during which they spoke primarily English, had a child who had clinical or borderline range scores on the Externalizing scale of the Child Behavior Checklist (CBCL) 1.5/5 (Achenbach & Rescorla, 2001) as reported by either the primary or alternate caregiver at the age 2 assessment. The mean age of the children was 29.5 months (*SD* = 3.2); 48% were female. Caregivers reported the children's ethnicities as: 53% European American, 27% African American, 7% Hispanic/Latino, 1% Native American/American Indian, and 12% biracial.

[1] Visual inspection indicated that, when compared to analyses run with all 79 cases, analyses run with the remaining 77 cases yielded miniscule differences.

Procedure

Home observation assessment protocol. The assessment, which occurred prior to randomization, was conducted as a 2.5-hour home visit with the target child and at least one caregiver (typically the mother). First, the child was given an assortment of toys to play with while the caregiver completed questionnaires. Next, the child and primary caregiver participated in a series of tasks: clean-up task (5 minutes), delay-of-gratification task (5 minutes), four teaching tasks (3 minutes each), free play (4 minutes), a second clean-up task (4 minutes), the presentation of two inhibition-inducing toys (2 minutes each), and a meal preparation and lunch task (20 minutes). This same home visit and observation protocol was followed at age 3. At each assessment, the assessment staff person rated caregiver involvement with and supervision of the target child. The staff person also videotaped interaction tasks, which were later rated by the FCU provider to be used for clinical purposes in the feedback session. These videotaped interactions were also rated by trained coders for use in analyses for scientific purposes. Assessments were conducted annually, with the average time between the age 2 and age 3 assessments 11.86 months (*SD* = 1.21).

Intervention conditions. Following the assessment, families were randomized to the intervention or control condition and those in the intervention condition were offered the FCU (Dishion & Stormshak, 2007), which involved a clinical interview followed by a session in which feedback was provided based on the family assessment. All the families in the current study participated in the FCU and participated in an age 2 feedback session conducted by one of thirteen providers.

Measures

COACH. Coders rated caregiver engagement and five aspects of fidelity.

Fidelity. Coders rated a recording of a feedback session to evaluate the extent to which the provider delivered key content, the skill with which intervention components were delivered, and the extent to which the delivery was consistent with the model and likely to promote behavior change. Prior to rating, coders reviewed a Feedback Form that contained information that the provider used to share assessment results with the families at the age 2 feedback sessions. Coders used the COACH rating form (Figure 1) while reviewing the session and consulted the COACH manual as needed. In addition to providing the descriptions below, the manual provides more detailed descriptions of the process skills associated with each domain and lists examples of accurate and inaccurate applications of these skills (e.g., an exemplar or non-exemplar statement or question on the part of the provider).

The dimensions rated for fidelity are as follows:

1. **C**onceptual accuracy and adherence to the FCU model. The provider demonstrates an accurate understanding of the FCU model in terms of its emphasis on family-centered change, caregiver leadership in the change process, and support of specific skills that define family management. The model is assessment driven and tailored to the specific needs of children and families; this unique aspect of the model shapes all provider–caregiver interactions.
2. **O**bservant and responsive to family's needs. It is essential that the provider observes the caregiver's immediate and pressing concerns and contextual factors

and responds accordingly while giving feedback or while working with the caregiver on changing a specific behavior. The delivery of feedback is appropriately modified to align with the caregiver's context and unique cultural and individual needs.

3. **A**ctive in structuring the session. The provider actively structures the change process using an assessment-driven case conceptualization as a guide. Aside from listening, being supportive, and being empathetic, the provider can use actions such as constructing useful questions, conducting role-plays, and redirecting discussions to motivate and empower the caregivers to behave differently in their interactions with children. The provider encourages caregiver involvement and uses active strategies to teach family management skills, which often require caregiver effort and self-regulation.
4. **C**areful when teaching. The foundation of the model is to use assessment data to direct the course of the family-centered intervention. Providers give feedback to caregivers to increase their accurate self-appraisals and motivation to either build on existing strengths or take corrective action in one or more areas. This dimension of the COACH evaluates whether the provider sensitively gives feedback and guidance to increase caregivers' motivation to change. Useful provider skills include reframes that incorporate family strengths, skillful questions that help caregivers reevaluate their motivations, and statements that validate the complexity of the change process.

5. **H**elpful in building hope and motivation. Specific therapeutic techniques from motivational interviewing are integrated into the FCU to promote caregivers' hope, motivation, and change. The motivational approach means (a) providing feedback to the caregiver, (b) acknowledging that the caregiver is responsible for the change process, (c) informing the caregiver about known effective change strategies, (d) providing the caregiver with a menu of change options and not controlling the change process by offering only one option, (e) expressing empathy for the caregiver's situation, and (f) promoting the caregiver's self-efficacy. These process skills are used in moment-by-moment interaction with caregivers to help the caregiver become an agent of positive family change and enhance motivation to work toward that end.

Each dimension was rated on a 9-point scale: *needs work* (1–3), *good work* (4 – 6), *exceptional work* (7–9). Scores in the "needs work" range indicate that the provider does not use the recommended skills and does not display a clear understanding of the principles of the model. Scores in the "good work" range indicate basic competence in the model, including an acceptable level of process skill and conceptual understanding of the model accompanied with occasional errors or missed opportunities. Scores in the "exceptional work" category indicate mastery of the process skills of the model and a clear understanding of its principles.

The COACH has demonstrated adequate reliability in previous studies, with ICCs of 0.52-0.76 on the individual dimensions (Smith et al., 2016; Smith et al., 2013) and high ICCs for the composite score at age 2 (0.74), age 3 (0.73), age 4 (0.77), and age 5

(0.71) (Chiapa et al., 2015). In an effectiveness study that included children ages 5 to 17 (M = 11.6, SD = 2.6), the ICC for the composite score was 0.73 (Smith, Stormshak, & Kavanagh, 2015). One study found that the COACH composite score predicted caregiver engagement, which predicted positive parenting one year later (Smith et al., 2013). Another study found that steeper declines in fidelity from ages 2 to 5 were related to less improvement in problem behaviors assessed at ages 7.5 and 8.5 (Chiapa et al., 2015).

Caregiver engagement during the feedback session. A 9-point scale was used to rate caregiver engagement. High Engagement (7-9) scores were given when a caregiver actively participated in the session by engaging in conversation and staying on topic, giving thoughtful responses, engaging in change talk, actively participating in role-play, asking questions, offering solutions, and showing initiative. Moderate Engagement (4-6) scores were given when a caregiver showed only modest or only occasional signs of engagement. Low Engagement (1-3) scores were given when a caregiver appeared disengaged or inattentive. Low scores indicate the caregivers did not attend to the provider, averted their gaze, demonstrated flat affect, displayed signs of boredom, gave very brief responses to the provider's questions, repeatedly got off topic, expressed doubt, or showed a lack of ability to process the provider's comments. Reliability of caregiver engagement has been good across previous studies (ICC: 0.80-0.87) (Dishion et al., 2014). Caregiver engagement has been shown to mediate the relation between fidelity and improvements in parenting and decreases in child problem behavior (Smith et al., 2013).

Coders and training process. Five coders rated fidelity and caregiver engagement of the 79 feedback sessions from Smith et al. (2013); two were later dropped from analyses because they were not video-recorded (only audio was available). The coders were undergraduate students with junior or senior standing recruited from upper-division psychology classes and received course credit for their participation as coders in the study.

The PI trained the coders. By the start of training, she had participated in research on the FCU for four years, was a certified FCU provider, and had conducted seven feedback sessions. She also helped to revise the COACH coding manual to increase clarity and usability, co-authored an e-learning course on the COACH, and coded three feedback sessions with 100% agreement (i.e., within one rating point on the dimension) with an FCU Implementation Scientist. During training and coding for the study, the PI attended COACH rating meetings with the FCU implementation team and consulted with members of that team when questions arose. The PI coded four sessions for reliability with an FCU Implementation Scientist during the training period and had the following percent agreement across the dimensions: 17%, 67%, 100%, and 83%. Average agreement between the PI and FCU Implementation Scientist across the seven sessions they double-coded was 81%.

Prior to training, the coders completed CITI certification in the following three courses: 1) IRB – Social & Behavioral Research, 2) RCR – Undergraduate Responsible Conduct of Research, and 3) RCR – Social and Behavioral Responsible Conduct of Research. Training began with four, hour-long meetings of a didactic nature. In the first

meeting, the PI reviewed confidentiality, gave an overview of the study, and assigned the FCU e-learning course as homework. In the second meeting, the coders and PI reviewed together one, 18-minute video that showed actors depicting excerpts from an actual feedback session. The PI paused the video every few minutes to discuss aspects of the protocol for a feedback session and clinical skills demonstrated by the provider. Prior to the third meeting, coders watched this video and three others that used actors (approximately 13 to 20 minutes in length). The coders were instructed to review handouts on motivational interviewing to help them "pick up on what the provider is doing or could be doing to motivate the caregivers to engage in the intervention process." They were asked to identify provider behaviors related to fidelity and caregiver behaviors indicating level of in-session engagement as demonstrated in the videos. These sessions were discussed in the third meeting, as were coders' answers to five multiple-choice questions about the FCU that assessed their ability to distinguish between high and low fidelity. Prior to the fourth meeting, coders completed the COACH e-learning module, which involved rating an actual feedback session and answering 11 multiple-choice questions about the COACH. In the fourth meeting, the PI provided feedback about how the coders' ratings of the feedback session compared to hers, which were considered the gold standard, and shared her observations of what the provider had done to demonstrate high fidelity and opportunities the provider had missed. This became the standard template for coding meetings, which began the fifth week of training. When a coder was unable to attend a meeting, the PI sent detailed written notes regarding the rationale for the gold standard scores on each dimension of the COACH.

During training, coders rated an average of two non-study sessions per week. The sessions used in training differed from study sessions in that most were of families who participated in the larger RCT with a child who did not have elevated scores on the Externalizing scale. Coders independently scored the sessions and submitted their ratings to the PI, who compared them to her own ratings to compute percent agreement. Acceptable reliability was defined as 83% (5 of 6 domains) agreement (defined as being within one rating point on the dimension) with the PI's scores. A total of 33 non-study sessions were discussed in the 17 weekly coding meetings. The first six meetings lasted one hour each, the following 11 lasted two hours.

During the first ten weeks of coding training, 19 sessions were rated using video review. In the tenth week, four of the five coders were reliable with the PI on both assigned sessions, all were reliable on at least one, and interrater agreement among the five coders was relatively high on both sessions (mean of agreement across all six dimensions for all rater dyads was 0.72 and 0.80). Training using audio review was then begun, lasting five weeks and including ten sessions. In the fourth week of audio training, three of the five coders were reliable with the PI on both sessions; for one of the sessions, all were reliable with the PI. For that session, the average interrater agreement among the coders was 0.97, whereas for the other session assigned that week, the average interrater agreement among the five coders was 0.63. Training was continued for another week. The four coders who rated sessions that week were all reliable with the PI on one of the two sessions assigned; only one of the four was reliable with the PI on the other. Training was continued using video review of four sessions. In the first week of this period, all

five coders were reliable with the PI on one session. For another session, only two coders were reliable with the PI but the average interrater agreement among the coders was 0.85. For a third session, two of the coders were reliable with the PI and the average interrater agreement among the coders was 0.55. The PI extended training another week. One session was assigned: none of the coders were reliable with the PI but interrater agreement among them was 0.87. The PI consulted with an FCU Implementation Scientist and decided – in part because of the high interrater reliability among the coders, and in part because of time constraints related to the approaching end of the semester – to begin coding the study sessions. At this point, each coder had, at some point in the training period, rated five consecutive sessions with at least 83% agreement with the PI's scores. Training is estimated to have taken approximately 90 hours.

During the period of rating study sessions, to minimize drift the PI met with the coders weekly as a group to compare their scores for one non-study session to those of the PI and discuss disagreements in coding. The team alternated each week between video and audio review of a unique non-study session (i.e., coders never reviewed the same session twice; this was also the case during initial training). In the fifth week of coding, none of the coders were reliable with the PI; percent agreement ranged from 17% to 67%. Coding was halted and another non-study session was assigned. Four of the five coders were reliable with the PI; the PI met individually with the fifth coder to discuss differences in ratings, and study coding was resumed. Across the seven non-study sessions coded to prevent drift, the percent of sessions for which coders had acceptable

reliability ranged from 29% to 57% (2 to 4 sessions). Average agreement ranged from 59% to 74%.

Coding. Each coder was assigned between 36 and 39 study sessions, half of them audio and half as video recordings. Each session was reviewed by 2 coders; as audio by 1 coder and as video by a different coder. Given five coders, there were ten unique pairs of coders (e.g., AB, AC, AD, etc.). To ensure equal distribution of rating modality across coders, coders within a pair alternated rating modalities. For the purpose of assigning both sessions and rating modalities, twenty unique coder-by-modality pairings were identified (e.g., AB, BA, AC, CA, etc.). Each pair rated 8 sessions, with each coder rating 4 sessions as audio and 4 as video (with one exception: as there were only 79 study sessions, one pair rated only 7 sessions). In addition to these assignments for creating primary ratings for each rating modality, 16 (20%) of the 79 sessions were assigned to be rated by a second pair of coders for the purpose of calculating interrater reliability for each rating modality. Assignment of double-coding was done in such a way that double-coding occurred throughout the study coding process (3 in the first week, 1 in the second week, 3 in the third week, 3 in the fourth week, 3 in the fifth week, and 3 in the sixth week). For double-coding, care was taken to include each of the 10 unique coder dyads in the set of 32 recordings that were double-coded, with dyads double-coding 2 to 4 sessions, and with 9 of the 10 dyads rating sessions using a mix of the two rating modalities. Once the pattern of coding assignments (see Figure 2) was created, as a list of sessions, one through 79, study sessions were randomly assigned to these coding slots by

generating a list of random numbers, one through 79, in Excel. The PI did not have access to the ratings of study sessions until after all ratings were completed.

Positive behavior support (PBS). The PBS construct is made of four observational measures, not all on the same scale, and was represented as a *z*-score (the PBS variable was the mean of the *z*-scores of each of the four measures). Three of the measures were scored by a team of trained undergraduates coders (different from those who coded fidelity for the current study) based on review of video recordings of the home observation assessment protocol. The fourth measure was scored by the person conducting the home visit. Additional information regarding coder training, reliability procedures, and the measurement model of PBS can be found in the parent study from which this subsample was drawn (Dishion et al., 2008; Lunkenheimer et al., 2008). The positive behavior support latent construct has been shown to be stable over time (Lunkenheimer et al., 2008) and to predict child problem behavior (Dishion et al., 2008), as well as inhibitory control and language skills (Lunkenheimer et al., 2008).

1. Behavior support. This measure is based on Relationship Process Code (RPC; (Jabson, Dishion, Gardner, & Burton, 2004) scores of positive reinforcement (verbal and physical), prompts and suggestions of positive activities, and positive structure (e.g., providing choices when asking for behavior change) from videotape coding (durations) of caregivers prompting and reinforcing the child's positive behavior. The kappa coefficient was 0.86 at both ages (Smith et al., 2013).
2. Engaged parent-child interactions. A measure of the average duration of parent-child sequences that involve talking or physical interactions (e.g., playing a game)

was created using RPC codes such as Talk and Neutral Physical Contact. Kappa coefficients were 0.86 at both ages (Smith et al., 2013).

3. Proactive parenting. Coders rated each caregiver on his or her tendency to anticipate potential problems and to provide prompts or other changes to avoid the child becoming upset or involved in problem behavior on the following items from the Coder Impressions Inventory (Dishion, Hogansen, Winter, & Jabson, 2004): parent gives child choices for behavior change whenever possible; parent communicates to the child in calm, simple, and clear terms; parent gives understandable, age-appropriate reasons for behavior change; parent adjusts/defines the situation to ensure the child's interest, success, and comfort; parent redirects the child to more appropriate behavior if the child is off task or misbehaves; parent uses verbal structuring to make the task manageable. The proactive parenting score is the mean of these six items. Cronbach's alpha for ages 2 and 3 were 0.84 and 0.87, respectively (Smith et al., 2013).
4. Parent involvement. The person conducting the home visit rated caregiver involvement using three items from the Home Observation for Measurement of the Environment (Bradley, Corwyn, McAdoo, & Coll, 2001): "Parent keeps child in visual range, looks at often"; "Parent talks to child while doing household work"; "Parent structures child's play periods" (Yes/No). The parent involvement score is the mean of those three items.

Data Analysis

Preliminary analyses included assessment of skewness and kurtosis for all study variables and exploratory factor analysis to determine the appropriateness of computing a composite fidelity score. ICCs for each dimension (five fidelity dimensions and caregiver engagement) for each rating modality were computed using raw scores and then using ratings standardized within coder and modality. Percent agreement between coders was also computed.

ICCs were tested across rating modalities for significant difference using a chi-square difference test. A paired *t*-test was conducted to test for mean differences across rating modalities for composite fidelity and caregiver engagement scores. Finally, a stacked model was used to test for differences in predictive validity across rating modalities. To address mono-rater bias in this study, analyses of predictive validity were conducted using the caregiver engagement scores rated by the current coders and again using caregiver engagement scores collected by a different team of coders and used in the analyses by Smith and colleagues (2013).

Sample statistics, correlations, and the paired *t*-test were computed using SPSS Version 24. All other analyses were carried out using path analysis in Mplus version 7.1 (Muthén & Muthén, 2013).

Results

Preliminary Analyses

Distributions of all variables were examined for non-normality. Skewness for each variable ranged from -0.46 to 0.02 and kurtosis ranged from -0.93 to 0.06 (see Table

1). These low absolute values, along with visual inspection of histograms, indicated normal distributions (Fidell & Tabachnick, 2003).

Composite measure of fidelity. Correlations between the five dimensions were computed. Pearson's *r* correlations ranging from 0.79 to 0.85 for ratings based on video review (Table 2) and from 0.82 to 0.91 for ratings based on audio review (Table 3) indicated the appropriateness of creating a composite score. This approach was verified using a principal axis factor analysis. The exploratory factor analyses that were conducted were limited to a maximum of two factors due to limited number of observed variables (Muthén & Muthén, 2009). Three statistics were examined to assess goodness-of-fit, using the following guidelines for good fit suggested by Hu and Bentler (1999): SRMR (< .08) and RMSEA (< .06), and CFI (> .95).

For ratings based on audio review, the two-factor solution would not converge. For the one-factor solution, the SRMR and CFI were consistent with good fit, whereas the RMSEA indicated adequate fit (SRMR .011; RMSEA .082; CFI .995). For ratings based on video review, the fit indices for the one-factor and two-factor solutions indicated good fit (SRMR .006; RMSEA <.001; CFI 1.000 and SRMR .001; RMSEA <.001; CFI 1.000, respectively). The one-factor solution compared against the two-factor solution had a chi-square value of 1.345, $df = 4$, $p = 0.854$. The one-factor did not fit significantly worse than the two-factor, consistent with the premise that the five domains all capture aspects of fidelity. These analyses indicated that both audio and video ratings of fidelity could be represented using a composite score. The five dimensions were equally weighted when creating the composite score.

Interrater Reliability

Sixteen of the 77 sessions were double-coded within rating modality. A one-way random-effects model intraclass correlation coefficient (ICC [1,1]; (Shrout & Fleiss, 1979) was computed for scores for each fidelity dimension, the composite fidelity score, and caregiver engagement separately for scores based on video review of sessions and audio review (see Table 4).

ICCs for each domain (including caregiver engagement), based on video review, ranged from 0.02 (hope and motivation) to 0.46 (caregiver engagement). For scores based on audio review, ICCs ranged from 0.01 (caregiver engagement) to 0.32 (active in structuring). The ICC for the composite fidelity score based on video review was 0.15; for audio, it was 0.12. With the exception of the ICC for caregiver engagement based on video review, which was fair, all other ICCs were poor, according to Cicchetti's guidelines for interpreting ICCs (1994).

Because the ICCs were lower than expected, additional analyses were conducted. ICCs were computed using fidelity scores standardized within rater and within rating modality to attempt to understand a potential cause of the low ICCs (i.e., perhaps coders had anchored differently, with a tendency toward higher or lower ratings). With the exception of the ICCs for active in structuring (0.47) and for caregiver engagement (0.56) based on video review, which were fair, all other ICCs were poor (i.e., less than 0.40).

Percent agreement was also examined to determine whether the coders' interrater agreement on study sessions was similar to what it had been during the training period. For study sessions, percent agreement for the dimensions of fidelity ranged from 56%

(careful teaching and hope and motivation) to 81% (active in structuring) and was 81% for caregiver engagement based on video review. For audio review, percent agreement for these dimensions of fidelity ranged from 63% (observant, careful, hope) to 69% (conceptually accurate, active in structuring) and was 63% for caregiver engagement. Mean agreement across the six dimensions coded using video was 70%; for audio, it was 65%. This was comparable to what it had been during the entire training period (65% agreement on average with the PI across the five coders and 66% agreement among the five coders). However, percent agreement on study sessions was somewhat lower than what it had been in the last eight training sessions (4 audio and 4 video) among the five coders (75%), although average agreement with the PI for those eight sessions was only 64%.

Comparison of ICCs for Audio and Video Ratings

To answer whether interrater reliability differed by rating modality, ICCs for each domain were constrained to be equal and a chi-square test was used to determine whether the ICCs for each rating modality were significantly different. There was not a significant difference for Conceptual accuracy (χ^2 (1) = .213, p = 0.64), Active in structuring (χ^2 (1) = .045, p = .83), or Careful when teaching (χ^2 (1) = .074, p = .79). The ICCs for Observant and responsive and Hope and motivation were equal, rendering a test for significant difference moot. A comparison of the ICC for the composite fidelity score indicated no significant difference (χ^2 (1) < .001, p > .999). For Caregiver engagement, there was not a significant difference (χ^2 (1) = 3.031, p = .08).

Correlations and Comparison of Audio and Video Ratings

Composite fidelity scores based on audio review were moderately correlated with those based on video ($r = .43$, $p < .001$, $n = 77$). The Pearson correlation coefficient for caregiver engagement scores were moderately correlated ($r = .43$, $p < .001$, $n = 77$).

There was not a significant difference in the composite fidelity scores based on audio review ($M = 5.45$, $SD = 1.64$) and video review ($M = 5.42$, $SD = 1.42$); t (76) = 0.18, $p = 0.86$, 95% CI = -0.34, 0.41. The computed mean difference score for composite fidelity was 0.0338 and Cohen's $d = 0.02$, indicating that the means for fidelity based on audio vs. video review differed by only 0.02 standard deviations. There was not a significant difference in ratings between modalities on any of the five dimensions of fidelity (see Table 5). For caregiver engagement based on audio review ($M = 6.26$, $SD = 1.48$) and video review ($M = 6.39$, $SD = 1.30$) there was not a significant difference; t (76) = -0.77, $p = 0.45$, 95% CI = -0.47, 0.21). The computed mean difference score for caregiver engagement was -0.14 and Cohen's $d = 0.09$, indicating that the means for caregiver engagement based on audio vs. video review differed by only 0.09 standard deviations.

Predictive Validity of COACH Measure for Audio vs. Video Ratings

First, in a path analysis framework and using maximum likelihood estimation, predictive validity was tested separately for audio and video ratings, using current ratings only and a combination of fidelity ratings from the current study and caregiver engagement ratings from the coders in Smith and colleagues' (2013) study. As shown in Figure 3, the path model included an indirect path from fidelity to positive behavior

support through caregiver engagement and a direct path from fidelity to positive behavior support, controlling for positive behavior support at baseline. The standardized path coefficients are presented in Tables 6-9 and in Figures 4-7. In all four models, the path from fidelity to caregiver engagement was positive and significant. The path from caregiver engagement to positive behavior support was positive and significant in the model using fidelity ratings based on audio review for the current study and caregiver engagement ratings from the coders in Smith and colleagues' (2013) study ($\beta = 0.223$, $p = 0.024$). The path from caregiver engagement to positive behavior support was not significant in the other models (fidelity and caregiver engagement ratings based on audio review: $\beta = 0.193$, $p = 0.162$; fidelity and caregiver engagement ratings based on video review by the current coders: $\beta = 0.128$, $p = 0.371$; fidelity based on video review by the current coders and caregiver engagement based on video review for Smith and colleagues' study: $\beta = 0.198$, $p = 0.051$). The direct path from fidelity to positive behavior support was not significant in any of the four models. The path from positive behavior support at baseline to positive behavior support one year later was positive and significant for all four models.

To compare the predictive validity of the rating modalities, a Wald Chi-Square test was used to examine the difference between individual parameters across groups in the model shown in Figure 3, using maximum likelihood estimation. Rating modality was used as the grouping variable in a two-group "stacked" model. Structural parameters were constrained to equality across the two modality groups and then allowed to vary. The null model was the model in which path coefficients were constrained to be equal

between the two models being "stacked" – the model for composite fidelity ratings based on video review and the model for fidelity ratings based on audio review – while the alternative model was such that path coefficients were allowed to differ. In this case, both models had indices consistent with good fit (for the constrained model, CFI >.999 and RMSEA <.001, 90% 1-RMSEA CI [0.000, 0.058]; for the unconstrained model, CFI >.999 and RMSEA <.001, 90% 1-RMSEA CI [0.000, 0.000]). The change in chi-square was nonsignificant ($\chi 2_{\text{diff}} = 1.977$, $df_{\text{diff}} = 3$, $p = .58$; see also Table 10). This nonsignificant change in chi-square indicates that the predictive validity of the measure did not significantly differ by rating modality. Because caregiver engagement and fidelity were measured by the same coder, which may have inflated the correlation between the two constructs, the model was also tested using caregiver engagement scores rated by the coders in Smith and colleagues' (2013) study. Again, both the constrained and unconstrained models had indices consistent with good fit (for the constrained model, CFI >.999 and RMSEA <.001, 90% 1-RMSEA CI [0.000, 0.060]; for the unconstrained model, CFI >.999 and RMSEA <.001, 90% 1-RMSEA CI [0.000, 0.091]). The change in chi-square was nonsignificant ($\chi 2_{\text{diff}} = 0.840$, $df_{\text{diff}} = 2$, $p = .66$; see also Table 11).

Discussion

This study compared the interrater reliability and predictive validity of coding of fidelity to the FCU, a family-centered, evidence-based preventive intervention using video recording versus audio recordings. There is a movement to advance the broad dissemination of prevention programs that have been demonstrated to be effective in controlled trials and to implement these programs in the most efficient and effective way

possible to maximize their public health benefits (Committee on the Prevention of Mental Disorders and Substance Abuse Among Children Youth and Young Adults: Research Advances and Promising Interventions, 2009). Given the consistent finding that fidelity to family-based interventions is significantly related to program outcomes (Forgatch & DeGarmo, 2011; Forgatch et al., 2005; Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Hogue et al., 2008; Huey Jr, Henggeler, Brondino, & Pickrel, 2000; Ogden & Hagen, 2008; Prado, Pantin, Schwartz, Lupei, & Szapocznik, 2005; Smith et al., 2013), an important issue for prevention science is the development of valid and reliable methods for assessing fidelity that are also feasible and acceptable.

Although there was not a significant difference in the reliability nor validity of fidelity scores based on audio versus video recording in the current study, caution must be used in interpreting these results because the interrater reliabilities in both conditions were low. With the exception of the ICC for caregiver engagement based on video review, which was fair, all other ICCs were poor. For video review, ICCS ranged from 0.02 to 0.46. For audio review, ICCs ranged from 0.01 to 0.32. The ICC for the composite fidelity score based on video review was 0.15; for audio, it was 0.12. Additionally, the study was underpowered. Below, possible explanations for the low interrater reliability are discussed, as are alternative methods for testing this research question and other directions for future research.

One factor that likely played a role in the low interrater reliability is related to the coders' background. They were undergraduate students who did not have a clinical background nor experience delivering the FCU. All five of the previous studies that

report acceptable interrater reliability on the COACH fidelity scores used coders who had experience delivering the FCU, with the exception of one coder who was on three of the four coding teams. Specifically, in the Smith et al. (2013) study, coders were two clinicians who were FCU providers and one undergraduate without clinical experience. The composite score ICC was 0.74. In the study conducted by Smith and other colleagues (2015), there were two coders, one a graduate psychology trainee and the same advanced undergraduate psychology student who coded in the Smith et al. (2013) study. The ICC for the composite score was 0.73. More recently, Smith and colleagues (2016) used three counseling psychology PhD students who had completed a 12-month practicum in the FCU, which required them to conduct two to five feedback sessions each (J.D. Smith, personal communication, May 5, 2018). For the condition in which coders reviewed assessment results prior to rating, ICC = .82. Chiapa and colleagues (2015) examined fidelity over time using data from the coders who had participated in the earlier studies by Smith and his colleagues (J.D. Smith, personal communication, May 5, 2018). The ICCs for the composite scores were 0.73, 0.77, and 0.71. Another study (Smith et al., under review) used ratings by two of the three counseling psychology doctoral students who were raters in the Smith et al. (2016) study; the non-clinician coder, now staff on the project, double-coded 20 percent of the sessions. For the study condition most similar to the current study, the ICC for the composite score was 0.82.

Given that all but one of the coders had experience delivering the FCU in the studies that report acceptable reliabilities, the current coders' lack of experience with the FCU may explain the low ICCs. The COACH is a complicated code intended to assess a

provider's flexible application of the principles underlying a model of intervention. Also, the assessment of fidelity entails attention to both content and process. It requires coders to make judgments about socially complex behaviors such as whether the provider gives feedback with empathy, demonstrates an accurate understanding of the FCU model, and observes and responds to the caregiver's immediate concerns and contextual factors. The current coders' lack of experience with the FCU and absence of any clinical background likely made it difficult for them to make these judgments reliably.

Another factor that may have contributed to the low reliability was the number of coders. The current study used five coders and computed ICCs across nine (video) or ten (audio) unique dyads, whereas the other studies that report reliability on COACH scores used two or three coders (Smith et al., 2016; Smith et al., 2013; Smith et al., under review; Smith et al., 2015). The COACH is a complex, global, subjective rating system that relies on clinical judgement and is applied to complex interpersonal interactions that last approximately an hour. Given this level of complexity, it may have been difficult to achieve consistently high levels of agreement across the five coders.

The nature of the training process may also have affected interrater reliability. As described in the methods section, prior to independently coding the training sessions, coders reviewed one training tape together and studied material on motivational interviewing. They also completed an e-learning course on the FCU, a quiz testing their ability to distinguish between high and low fidelity behaviors, and the COACH e-learning module, which involved rating fidelity of an actual feedback session. Coders then scored multiple non-study sessions and met as a group to discuss their scores.

During these meetings, the PI provided a rationale for the gold standard scores. The total training time was about 90 hours.

Details are not available regarding the nature of the training process for the Smith et al. (2013) study. All that is known is that the program developer shared training responsibilities with a BA-level study coordinator who had assisted in writing the COACH manual (J.D. Smith, personal communication, May 5, 2018). For the Smith et al. (2016) study, the coders and PI together scored two tapes that had high interrater agreement when double-coded by a previous coding team, pausing every two to five minutes to discuss what they observed. The coders then rated a session independently and reviewed it together. After independently scoring and getting feedback on their scores, the coders rated two sessions independently and met criterion for mastery across the three sessions. In these studies, training took approximately 20 hours (Smith et al., 2016; Smith et al., 2013). The Chiapa et al. (2015) study, the Smith et al. (2015) study, and the Smith et al. study (under review) used only coders from previous coding teams, making additional training unnecessary (J.D. Smith, personal communication, May 30, 2018).

The primary difference in the training process for the previous COACH coding team for which information is available (Smith et al., 2016) is that the PI reviewed multiple complete feedback sessions with the coders, stopping every few minutes for discussion. Training for this study may have been improved by using the same approach, for all dimensions at once or with a focus on one dimension at a time. Reviewing sessions together, with the PI narrating her rating process throughout, could have led to the coders better understanding the appropriate use of the rating dimensions. Another process that

may have led to higher reliability would have been to have coders write their definition of each dimension, with exemplars and non-exemplars, so the PI could address both common and unique misunderstandings.

It is also possible that the low reliability was in part due to the criterion for beginning coding of study sessions. Although in the current study it was planned that the coders would all have coded the same three consecutive sessions at 83% agreement with the gold standard, it was decided to have the coders rate the study sessions after about 90 hours of training even though mastery of the system was not consistently demonstrated. By that time each coder had, at some point in the training period, rated five consecutive sessions with at least 83% agreement with the PI's scores. However, while this level of agreement occurred at some point in training, it was not necessarily immediately before study coding commenced, nor were all coders at this level right before study coding began. In contrast, mastery of the fidelity rating process, as demonstrated by at least 85% agreement across the 18 domains rated across three training sessions, was achieved by the coders in other studies that report acceptable reliability of the COACH (Smith et al., 2016; Smith et al., 2013).

A fifth factor that could be related to the low reliability is the short period in which coders rated the sessions. Because training took so much longer than anticipated and the coders, who were students receiving course credit, were available only through the end of the school year, the coders rated six study sessions and one non-study session each week, which may have overburdened them. The study sessions averaged 67 minutes (*SD* = 19.3). Thus, each coder completed approximately eight hours of coding each week

for at least six weeks. The need to code seven sessions each week, combined with the high demand for vigilance while coding, may have led to poorer quality of coding and thus to low reliability. The timeline for coding for other COACH studies is unreported, so it is impossible to compare this aspect of the coding process with those in other studies.

There are some other factors that were considered as possible reasons for the differences in reliability between this study and the previous studies on the COACH. Although the nature of the coding meetings during coding of the study sessions in the current study and those in the previous studies differed, it seems unlikely that difference was responsible for the differences in reliability. In the current study, coders reviewed one non-study session per week and met weekly to discuss the gold standard scores for that session. In other words, training was continued during the coding of the study sessions, as coders had not demonstrated consistent high agreement with the PI's ratings when coding began. Smith and colleagues (2013) reported that the coding team met weekly to prevent drift; the process for those meetings was not reported. Smith and colleagues (2016) reported that the coding team met biweekly to discuss the coding challenges they had encountered over the past two weeks while coding study sessions (J.D. Smith, personal communication, May 5, 2018). The coding team for another study by Smith and colleagues (2015) did not meet during coding of the study sessions, perhaps because they coded just 32 sessions among two coders (i.e., likely only 16 sessions each). Given that average agreement with gold standard scores on non-study sessions was 64% on the final eight training sessions and 66% during coding of the study sessions, the procedure used in the current study appears to have prevented drift. It is much more

likely that the low ICCs are attributable to the fact that the coders never reached consistent reliability during the training period, and reliability did not substantially improve during the study coding period.

Recording quality was also considered as a potential influence on reliability. For six of the sixteen double-coded sessions, coders reported difficulty understanding what was being said due to ambient noise (e.g., loud toy or the sound of a fan). However, the average difference in composite scores between coders was not greater for recordings reported to be of poor quality than for recordings in which problems were not reported, which suggests that recording quality did not contribute to low reliability.

In summary, the factors most likely contributing to the low interrater reliability include the current coders' lack of experience with the FCU and absence of any clinical background, the training process, and commencement of study coding prior to consistent demonstration of mastery. The relatively large number of coder dyads, given the complexity of the COACH, and the short timeline for coding may also have contributed.

Because interrater reliability in both audio and video conditions was low, findings are inconclusive regarding whether these rating modalities yield scores that are equally reliable and valid. However, this research question is important to examine in future research. The FCU is currently in the process of scale-up and is being disseminated using the FCU Implementation Framework (IF) to support implementation with quality (Mauricio, Rudo-Stern, Dishion, Letham, et al., under review). FCU consultants use the COACH to measure fidelity during the certification process, and FCU supervisors at

implementation sites are trained to use the COACH for ongoing fidelity monitoring to sustain high quality implementation.

Attention to fidelity during training and program delivery is important given that fidelity is significantly and directly related to client engagement, which predicts later parenting and child behavior problems (Smith et al., 2013), and declines in fidelity are related to less improvement in child problem behaviors (Chiapa et al., 2015). Further, drift in fidelity does occur but declines can be prevented when fidelity checks are implemented and issues are addressed in clinical supervision (Chiapa et al., 2015). Implementation sites have expressed concern, however, regarding the intrusiveness of video recording, related privacy issues, and the time required to upload large video data files to a secure server. One implementation site insisted on using audio recording due to these concerns. Further, the question of whether audio and video rating modalities are comparable is relevant to the development of pragmatic measures for monitoring fidelity for other parenting-focused, evidence-based programs that are increasingly being disseminated in community settings.

Limitations

As discussed above, reliability was low, which limits the ability to draw conclusions about the equivalence of audio and video modalities for rating fidelity and the predictive validity of the COACH based on audio versus video review. Furthermore, the sample was small, which limited the power to detect differences in interrater reliability and predictive validity across the modalities and would have been a limitation even if reliability had been high. Also, there are several aspects of the sample and

procedures that limit the generalizability of the findings. First, the study included only families in which the target child was 2 years old and had elevated problem behaviors. Second, all but six feedback sessions rated for fidelity in this study were conducted as home visits. Third, only feedback sessions from an efficacy trial were included. Using recordings from multiple waves of the RCT, and from a larger sample of families in the RCT, as well as recordings from the effectiveness trial would have increased variation in child age, child clinical presentation, and service delivery settings. Furthermore, the current study compared audio and video rating modalities using only the COACH fidelity rating system, which limits the generalizability of its findings.

Future directions

One direction for future research would be to address this question using coders who are FCU providers. The training would be modified to include group review and moment-by-moment discussion of complete, actual feedback sessions, as well as written assignments to assess understanding of the rating dimensions. Study coding would not begin until all coders reach 83% agreement on each of the same three consecutive sessions. Ideally, the recordings would come from in-clinic sessions, to capitalize on higher audio quality and to reflect increasing FCU implementation in clinics. To have adequate power to detect significant differences in reliability and validity across rating modality, a larger number of sessions would be coded.

Additionally, future research might address ways to simplify the COACH rating process to create a more acceptable and feasible system for rating fidelity. There is currently a disconnect between how fidelity is measured for research (i.e., using a 9-point

scale) and how it is used in training and certification. Although consultants use the 9-point scale to evaluate the providers' performance during training and supervision, only qualitative data related to the COACH criteria, rather than specific scores, are typically discussed with providers. For certification, a threshold for adequate versus inadequate fidelity (i.e., threshold of 4 or higher on all five dimensions) is used, and variance above or below a score of 4 is irrelevant. As an alternative to using the 9-point scale during training and supervision, consultants could rate simply whether the provider demonstrated adequate fidelity. The challenges that FCU consultants and supervisors have experienced in achieving and maintaining reliability using the 9-point scale argues for the exploration of simplifying the fidelity-rating process for the purposes of certification, supervision, and ongoing fidelity monitoring. Such a solution may not work for research purposes, due to the restriction of variance, but there may be an alternative method for fidelity rating for research that could better align how the COACH is used in research with how it is used in clinical practice.

Another way to increase the feasibility of the COACH would be to review only a segment instead of a complete session. Smith and colleagues (under review) compared coding of 20-minute segments to complete sessions drawn from an efficacy trial and from an effectiveness trial. Although the magnitude of composite scores did not differ significantly between ratings of segments versus ratings of complete sessions, ICCs for the ratings of segments were significantly lower than those for complete sessions in the effectiveness trial (i.e., providers had a level of training and supervision typical of providers at implementation sites). ICCs for ratings of segments versus complete sessions

in the efficacy trial were excellent and did not significantly differ. These findings suggest that ongoing monitoring of highly trained, experienced, and skilled providers who are vulnerable to drift (Chiapa et al., 2015) could be done by rating segments, whereas complete sessions would be need to be rated for providers newer to the model.

It is also possible that a different selection process could be used to identify segments to review for providers with less experience delivering the FCU. For example, 10-minute segments might be chosen from the beginning, middle, and end of the session, so as to capture particular elements of the intervention (i.e., explanation of the process, discussion of assessment results, and goal-setting). Another, similar fidelity rating system uses 10-minute segments selected by a trained assistant to capture important aspects of the intervention (e.g., debriefing home practice, role playing, problem-solving) and has achieved good to excellent interrater reliability (Forgatch et al., 2005). Rather than have a trained assistant skim recordings to identify the most salient segments, the provider who conducted the session could identify the segments that include essential components of the intervention. Such segments could be reviewed for the purposes of ongoing fidelity monitoring and clinical supervision. Should the ratings from selected segments be reliable, it would be important to examine the relation between these scores and client outcomes.

Conclusions

Given that fidelity to evidence-based parenting programs is significantly related to client outcomes (Forgatch & DeGarmo, 2011; Forgatch et al., 2005; Gillespie et al., 2017; Henggeler et al., 1997; Hogue et al., 2008; Huey Jr et al., 2000; Ogden & Hagen,

2008; Prado et al., 2005; Smith et al., 2013), for the broad dissemination of these programs to have a positive impact on public health they must be delivered in community settings with high fidelity (Glasgow, Klesges, Dzewaltowski, Bull, & Estabrooks, 2004; Spoth et al., 2013). To achieve and sustain a high level of fidelity, continued fidelity monitoring and consultation may be required (Schoenwald, Sheidow, & Letourneau, 2004). For this to occur, reliable and valid systems for monitoring fidelity and providing remediation that are also feasible and acceptable are needed. The review of audio recordings may be a more feasible and acceptable alternative to video for rating fidelity. Although this study's findings are inconclusive, due to low interrater reliability and statistical power, the question of whether ratings of audio versus video recordings are comparable is one with important public health implications. Were audio determined to be as reliable and valid for rating fidelity, the FCU and other evidence-based programs could adapt their implementation models to replace video with audio review. Such a change might increase the acceptability and feasibility of the evidence-based, family-centered programs themselves, resulting in greater adoption and, in turn, reach. This change might also increase the likelihood of ongoing fidelity monitoring and consultation, which could improve the quality of services families receive. Further research is needed to test the equivalence of audio and video modalities for rating fidelity, with the potential to increase the public health impact of evidence-based, family-centered programs.

REFERENCES

Achenbach, T. M., & Rescorla, L. (2001). *ASEBA school-age forms & profiles*: Aseba Burlington.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*(1), 23-33.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Coll, C. G. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844-1867.

Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research, 59*(3), 158.

Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nursing & Health, 33*(2), 164-173.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40.

Chiapa, A., Smith, J. D., Kim, H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2015). The trajectory of fidelity in a multiyear trial of the Family Check-Up predicts change in child problem behavior. *Journal of Consulting and Clinical Psychology, 83*(5), 1006.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284.

Committee on the Prevention of Mental Disorders and Substance Abuse Among Children Youth and Young Adults: Research Advances and Promising Interventions. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities* (M. E. O'Connell, T. Boat, & K. E. Warner Eds.). Washington, DC: The National Academies Press.

Dent, E., Brown, R., Dowsett, S., Tattersall, M., & Butow, P. (2005). The Cancode interaction analysis system in the oncological setting: reliability and validity of video and audio tape coding. *Patient education and counseling, 56*(1), 35-44.

Dishion, T. J., Connell, A., Weaver, C., Shaw, D., Gardner, F., & Wilson, M. (2008). The Family Check-Up with high-risk indigent families: Preventing problem behavior

by increasing parents' positive behavior support in early childhood. *Child Development, 79*(5), 1395-1414.

Dishion, T. J., Hogansen, J., Winter, C., & Jabson, J. (2004). The Coder Impressions Inventory. Unpublished manuscript, Child and Family Center; Eugene, OR.

Dishion, T. J., Smith, J. D., Gill, A., Shaw, D., & Knutson, N. (2014). Family Check-Up & Everyday Parenting (EDP): Fidelity COACH Rating Manual: Version 4.0. Unpublished coding manual. Available from the REACH Institute at Arizona State University.

Dishion, T. J., & Stormshak, E. A. (2007). *Intervening in children's lives: An ecological, family-centered approach to mental health care*. Washington, DC: American Psychological Association.

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Phillips Smith, E., & Prinz, R. J. (2001). Promoting intervention fidelity. *American Journal of Preventive Medicine, 20*(1), 38-47. doi:10.1016/S0749-3797(00)00272-5

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327-350.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237-256.

Dusenbury, L., Brannigan, R., Falco, M., & Lake, A. (2004). An exploration of fidelity of implementation in drug abuse prevention among five professional groups. *Journal of Alcohol and Drug Education, 47*(3), 4.

Eames, C., Daley, D., Hutchings, J., Hughes, J. C., Jones, K., Martin, P., & Bywater, T. (2008). The Leader Observation Tool: a process skills treatment fidelity measure for the Incredible Years parenting programme. *Child: Care, Health and Development, 34*(3), 391-400. doi:10.1111/j.1365-2214.2008.00828.x

Eeckhout, T., Gerits, M., Bouquillon, D., & Schoenmakers, B. (2016). Video training with peer feedback in real-time consultation: Acceptability and feasibility in a general-practice setting. *Postgraduate Medical Journal*. doi:10.1136/postgradmedj-2015-133633

Fidell, L. S., & Tabachnick, B. G. (2003). Preparatory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of Psychology* (Vol. 2, pp. 115-142). Hoboken, NJ: Wiley.

Forgatch, M. S., & DeGarmo, D. S. (2011). Sustaining fidelity following the nationwide PMTO™ implementation in Norway. *Prevention Science, 12*(3), 235-246.

Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior therapy, 36*(1), 3-13.

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review, 31*(1), 79-88.

Gillespie, M. L., Huey, S. J., & Cunningham, P. B. (2017). Predictive validity of an observer-rated adherence protocol for multisystemic therapy with juvenile drug offenders. *Journal of Substance Abuse Treatment, 76*, 1-10.

Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine, 27*(1), 3-12.

Haine-Schlagel, R., & Walsh, N. E. (2015). A review of parent participation engagement in child and family mental health treatment. *Clinical Child and Family Psychology Review, 18*(2), 133-150.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*(5), 821-833.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*(4), 544-555. doi:10.1037/0022-006x.76.4.544

Huey Jr, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*(3), 451-467.

Jabson, J., Dishion, T. J., Gardner, F., & Burton, J. (2004). Relationship Process Code v-2.0. training manual: A system for coding relationship interactions. *Child and Family Center, 160*.

Lunkenheimer, E. S., Dishion, T. J., Shaw, D. S., Connell, A. M., Gardner, F., Wilson, M. N., & Skuban, E. M. (2008). Collateral benefits of the Family Check-Up on

early childhood school readiness: Indirect effects of parents' positive behavior support. *Developmental Psychology, 44*(6), 1737-1752.

Mauricio, A. M., Rudo-Stern, J., Chiapa, A., Smith, J. D., Dishion, T. J., & Berkel, C. (under review). A dynamic, data-based decision-making system to assess and build readiness to implement the Family Check-Up.

Mauricio, A. M., Rudo-Stern, J., Dishion, T. J., Letham, K., & Lopez, M. (under review). Data-driven adaptation of implementation components during scale-up of the Family Check-Up.

Mauricio, A. M., Rudo-Stern, J., Dishion, T. J., Shaw, D., Gill, A., Lundgren, J., . . . the Center for Progress in Children's Mental Health. (under review). Understanding the cross-cultural transportability of evidence-based parenting programs by exploring implementation facilitators and barriers: A case study using the Family Check-Up.

Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth, 4*(4), 83-105.

Muthén, L. K., & Muthén, B. O. (2009). Mplus Short Courses Topic 1: Exploratory Factor Analysis, Confirmatory Factor Analysis, And Structural Equation Modeling For Continuous Outcomes.

Muthén, L. K., & Muthén, B. O. (2013). Mplus (Version 7.1). Los Angeles, CA.

Ogden, T., & Hagen, K. A. (2008). Treatment effectiveness of Parent Management Training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology, 76*(4), 607-621. doi:10.1037/0022-006X.76.4.607

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*(4), 365-383.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007, Dec 2007). *Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors.* Paper presented at the Annual Convention of the Association for Psychological Science, 18th, May, 2006, New York, NY, US; This work was presented in part at the aforementioned conference.

Prado, G., Pantin, H., Schwartz, S. J., Lupei, N. S., & Szapocznik, J. (2005). Predictors of engagement and retention into a parent-centered, ecodevelopmental HIV preventive intervention for Hispanic adolescents and their families. *Journal of Pediatric Psychology, 31*(9), 874-890.

Riddle, D. L., Albrecht, T. L., Coovert, M. D., Penner, L. A., Ruckdeschel, J. C., Blanchard, C. G., . . . Urbizu, D. (2002). Differences in audiotaped versus videotaped physician-patient interactions. *Journal of Nonverbal Behavior, 26*(4), 219-239.

Sanders, M. R., & Kirby, J. N. (2012). Consumer engagement and the development, evaluation, and dissemination of evidence-based parenting programs. *Behavior Therapy, 43*(2), 236-250.

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*(1), 94-104.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

Smith, J. D., Dishion, T. J., Brown, K., Ramos, K., Knoble, N. B., Shaw, D. S., & Wilson, M. N. (2016). An experimental study of procedures to enhance ratings of fidelity to an evidence-based family intervention. *Prevention Science, 17*(1), 62-70.

Smith, J. D., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2013). Indirect effects of fidelity to the Family Check-Up on changes in parenting and early childhood problem behaviors. *Journal of Consulting and Clinical Psychology, 81*(6), 962-974. doi:10.1037/a0033950

Smith, J. D., Rudo-Stern, J., Dishion, T. J., Stormshak, E. A., Montag, S., Brown, K., . . . Wilson, M. N. (under review). A quasi-experimental study of the effectiveness and efficiency of observationally assessing fidelity to a family-centered intervention.

Smith, J. D., Stormshak, E. A., & Kavanagh, K. (2015). Results of a Pragmatic Effectiveness–Implementation Hybrid Trial of the Family Check-Up in Community Mental Health Agencies. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(3), 265-278.

Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., . . . Hawkins, J. D. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science, 14*(4), 319-351.

Weingarten, M. A., Yaphe, J., Blumenthal, D., Oren, M., & Margalit, A. (2001). A comparison of videotape and audiotape assessment of patient-centredness in family physicians' consultations. *Patient education and counseling, 45*(2), 107-110.

Williams, K., Herman, R., & Bontempo, D. (2013). Comparing audio and video data for rating communication. *Western Journal of Nursing Research, 35*(8), 1060-1073.

Table 1

Sample characteristics

		Audio							Video						
	Variable	n	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis	n	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis
Ratings from 2017 Coding Team	Conceptually Accurate	77	2	9	5.56	1.68	-0.05	-0.66	77	2	9	5.57	1.40	-0.08	0.06
	Observant and Responsive	77	2	9	5.43	1.71	<0.01*	-0.61	77	1	9	5.52	1.64	-0.01	-0.33
	Active in Structuring	77	2	9	5.48	1.74	0.02	-0.93	77	2	8	5.27	1.43	-0.05	-0.64
	Careful when Teaching	77	1	8	5.29	1.77	-0.16	-0.74	77	1	8	5.35	1.71	-0.19	-0.41
	Hope and Motivation	77	2	9	5.51	1.78	-0.02	-0.70	77	2	8	5.38	1.51	-0.29	-0.58
	Composite COACH Score	77	2.00	8.60	5.45	1.64	-0.03	-0.85	77	1.60	8.40	5.42	1.42	-0.22	-0.38
	Caregiver Engagement	77	3	9	6.26	1.48	-0.46	-0.34	77	3	9	6.39	1.30	-0.29	-0.44
Ratings from 2013 Coding Data	Caregiver Engagement								77	3	9	5.99	1.60	-0.24	-0.87
Parenting Data	Positive Behavior Support Age 2								77	-1.54	1.04	-0.06	0.56	-0.43	0.25
	Positive Behavior Support Age 3								77	-1.32	1.61	0.12	0.63	-0.14	-0.43

Note. Skewness Std. Error of 0.27; Kurtosis Std. Error of 0.54. *Skewness for Observant and Responsive was -0.004.

Table 2

Zero-order correlations, video

	1	2	3	4	5	6	7	8	9
1 Conceptual accuracy and adherence to the FCU model	—	.79**	.80**	.82**	.80**	.91**	.61**	.17	.23*
2 Observant and responsive to client needs		—	.81**	.85**	.82**	.93**	.70**	.15	.26*
3 Accurately structures sessions to optimize effectiveness			—	.82**	.80**	.91**	.63**	.10	.18
4 Careful and appropriate teaching and corrective feedback				—	.82**	.94**	.67**	.22	.29*
5 Hope and motivation are generated					—	.92**	.73**	.14	.18
6 COACH composite score						—	.72**	.17	.25*
7 Observed client engagement in the feedback session							—	.12	.24*
8 Positive behavior support (child age 2)								—	.45**
9 Positive behavior support (child age 3)									—

Note. $N = 77$. Correlations were calculated using a Pearson's *r*. FCU = Family Check-Up.

$^{*}p < .05$. $^{**}p < .01$.

Table 3

Zero-order correlations, audio

	1	2	3	4	5	6	7	8	9
1 Conceptual accuracy and adherence to the FCU model	—	.83**	.88**	.91**	.84**	.95**	.62**	-.03	.15
2 Observant and responsive to client needs		—	.82**	.84**	.85**	.92**	.64**	-.10	.11
3 Accurately structures sessions to optimize effectiveness			—	.90**	.84**	.94**	.63**	-.09	.06
4 Careful and appropriate teaching and corrective feedback				—	.89**	.96**	.70**	-.01	.04
5 Hope and motivation are generated					—	.94**	.66**	.05	.07
6 COACH composite score						—	.69**	-.04	.09
7 Observed client engagement in the feedback session							—	<.01	.17
8 Positive behavior support (child age 2)								—	.45**
9 Positive behavior support (child age 3)									—

Note. $N = 77$. Correlations were calculated using a Pearson's *r*. FCU = Family Check-Up.

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

Table 4

Interrater reliability

	C	O	A	C	H	Composite	Engagement
Video ICC	.19	.17	.39	.04	.02	.15	.46
Audio ICC	.04	.17	.32	.13	.02	.12	.01
Video z-scores ICC	.14	.31	.47	.12	.11	.28	.56
Audio z-scores ICC	.19	.31	.37	.37	.11	.29	.14
Video % agreement +/-1	.75	.69	.81	.56	.56		.81
Audio % agreement +/-1	.69	.63	.69	.63	.63		.63

Note. Percent agreement is not presented for the composite fidelity score because the composite is a computed mean of coders' ratings on the five dimensions, not something coders rated directly. Overall agreement across the five dimensions of fidelity based on video review would be 67%; 65% for audio.

Table 5

Paired Samples Test: Paired Differences

Audio - Video	Mean	SD	95% CI of the Difference	*t*	df	*p* value
Conceptual	-0.013	1.80	-0.420, 0.395	-0.063	76	0.950
Observant	-0.091	1.89	-0.519, 0.337	-0.423	76	0.674
Active	0.208	1.71	-0.181, 0.596	1.065	76	0.290
Careful	-0.065	1.93	-0.503, 0.373	-0.295	76	0.768
Hope	0.130	1.85	-0.291, 0.550	0.615	76	0.540

Table 6

Path coefficients of the model, all ratings from current study, audio review

	Estimate (standardized)	S.E.	Est./S.E.	P-Value
PBS age 3 ON				
PBS age 2	0.450	0.090	4.994	<0.001
Fidelity	-0.030	0.139	-0.219	0.827
Caregiver Engagement	0.193	0.138	1.399	0.162
Caregiver Engagement ON				
Fidelity	0.691	0.059	11.626	<0.001

Note. $N = 77$. PBS = positive behavior support construct.

Table 7

Path coefficients of the model, Caregiver Engagement from Smith et al., 2013, fidelity based on audio review

	Estimate (standardized)	S.E.	Est./S.E.	P-Value
PBS age 3 ON				
PBS age 2	0.461	0.087	5.295	<0.001
Fidelity	0.046	0.101	0.455	0.649
Caregiver Engagement	0.223	0.099	2.259	0.024
Caregiver Engagement ON				
Fidelity	0.255	0.107	2.393	0.017

Note. $N = 77$. PBS = positive behavior support construct.

Table 8

Path coefficients of the model, all ratings from current study, video review

	Estimate (standardized)	S.E.	Est./S.E.	P-Value
PBS age 3 ON				
PBS age 2	0.420	0.092	4.560	<0.001
Fidelity	0.084	0.145	0.579	0.563
Caregiver Engagement	0.128	0.143	0.895	0.371
Caregiver Engagement ON				
Fidelity	0.723	0.054	13.310	<0.001

Note. $N = 77$. PBS = positive behavior support construct.

Table 9

Path coefficients of the model, Caregiver Engagement from Smith et al., 2013, fidelity from current study, video review

	Estimate (standardized)	S.E.	Est./S.E.	P-Value
PBS age 3 ON				
PBS age 2	0.438	0.089	4.928	<0.001
Fidelity	0.105	0.105	1.002	0.316
Caregiver Engagement	0.198	0.101	1.955	0.051
Caregiver Engagement ON				
Fidelity	0.336	0.101	3.326	0.001

Note. $N = 77$. PBS = positive behavior support construct.

Table 10

Stacked path model with engagement scores rated for current analyses

Multiple group model	**df**	**CFI**	**RMSEA Est/CI**	**χ2**	**χ2 Difference**	**df Difference**	***p* value**
Constrained	6	1.00	0.000/0.000-0.058	2.109	1.977	3	0.58
Unconstrained	3	1.00	0.000/0.000-0.000	0.132			

Table 11

Stacked path model with engagement scores from Smith et al., 2013

Multiple group model	**df**	**CFI**	**RMSEA Est/CI**	**χ2**	**χ2 Difference**	**df Difference**	***p* value**
Constrained	6	1.00	0.000/0.000-0.060	2.134	0.840	2	0.66
Unconstrained	4	1.00	0.000/0.000-0.091	1.294			

FCU Feedback Session COACH Rating Form

Version 4.3

Exceptional work	Competent work	Needs work
9 8 7	6 5 4	3 2 1

Provider:_____ Family ID:_____ TC Age:_____ Date of Feedback Session:_____ Rater:_____

Conceptually accurate and adherent to the model	**Barriers to effective practice**	
☐ Follows FCU protocol and principles in structure and content of session ☐ Gives data-based feedback using the Feedback Form and video of FIT ☐ Prioritizes and frames strengths/areas of concern using a parenting focus ☐ Offers a menu of evidence-based services that address family's specific needs	☐ Avoids feedback or minimizes areas that need attention ☐ Delves into tangents or engages in speculations that are NOT evidence-based ☐ Shows a premature focus on blaming or on family's stories that detract from FCU	☐
Observant and responsive to family's needs		
☐ Establishes a collaborative set using reflective listening and empathy ☐ Tailors feedback to caregiver's education, emotional needs, and cultural background ☐ Uses language and examples that are those of the caregiver and reflect the family storyline and social context ☐ Incorporates caregiver's immediate concerns and context and adjusts session agenda and intervention methods accordingly	☐ Isn't responsive to caregiver's input or behavior in session ☐ Lectures, steam rolls; disproportionate therapist/client talk ratio (goal is 1:1) ☐ Misses potential issues related to harm reduction or immediate action	☐
Active in structuring the session		
☐ Begins with caregiver self-assessment, explains Feedback Form, gives feedback, summarizes key points, and discusses follow-up services ☐ Manages discussion to weave in suggestions on intervention options ☐ Is prepared and uses materials appropriately (e.g., video feedback, handouts) ☐ Asks for caregiver's perspective; invites contributions and responses	☐ Mismanages time; session too long or significant sections of session unrealized ☐ Session disjointed; needs better pacing, transitions, and time for instruction ☐ Session structure disrupts client's potential to 'get' the feedback and respond	☐
Careful when teaching and providing feedback		
☐ Identifies and builds on existing strengths ☐ Tailors and scaffolds feedback and support to caregiver's abilities ☐ Connects the dots between the family's earlier reports and the feedback ☐ Uses research-based rationales and evidence-based intervention procedures ☐ Provides video feedback on caregiver-child interaction	☐ Minimizes or avoids areas of concern/opportunities for feedback and correction ☐ Provides rationales or advice that is unscientific or unprofessional ☐ Teaches too much in feedback session (i.e., lecturing, too much information)	☐
Helpful in building hope and motivation		
☐ Offers validation, empathy, and hopeful reframing ☐ Prompts, evokes, and supports change talk ☐ Instills hope by identifying strengths and reflecting on previous successes ☐ Supports self-efficacy by identifying realistic goals with achievable steps	☐ Misses opportunities to highlight past client efforts, successes, or strengths ☐ Advice giving, disagreement, or teaching in the face of ambivalence or discord ☐ Either through words or actions, gives discouraging message that undermines change (e.g., blame, critical comment, etc.)	☐
Caregiver engagement		
☐ Actively participates, nods head, and stays on topic during feedback ☐ Gives thoughtful responses to therapist's questions; demonstrates understanding ☐ Engages in "change talk" by reflecting on the past and future ☐ Articulates problems, goals; wishes to do things differently	☐ Angry or defensive during feedback session ☐ Does not share information and is not open about family life ☐ Seems unconcerned about parenting and/or child	CG1 CG2 ☐ ☐

Technical difficulties/notes:

Video tape used: Y N

Figure 1. COACH rating form v 4.3.

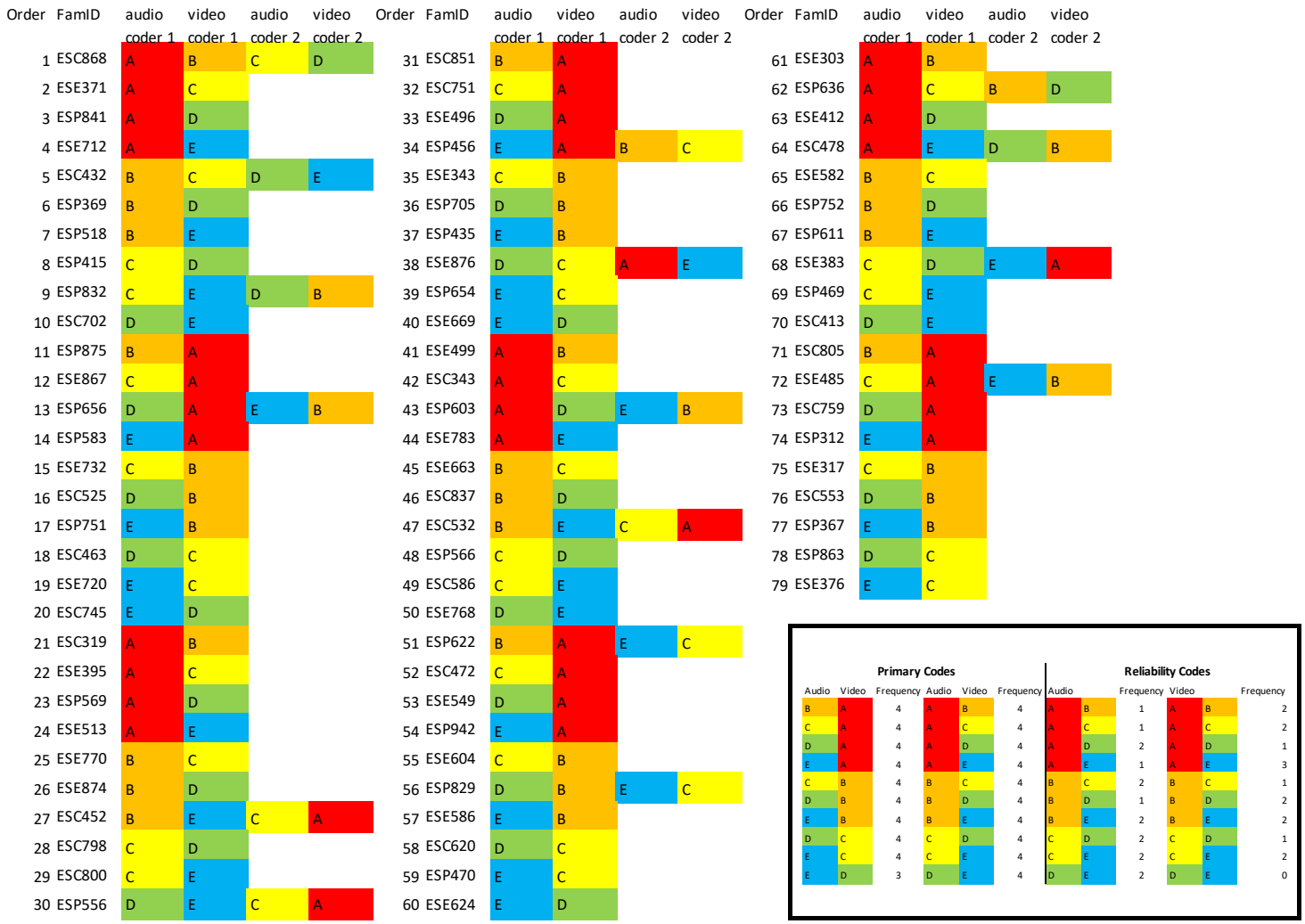

Figure 2. Coding assignments. Sessions ESP415 and ESP470 were later dropped from the analyses due to being recorded as audio-only.

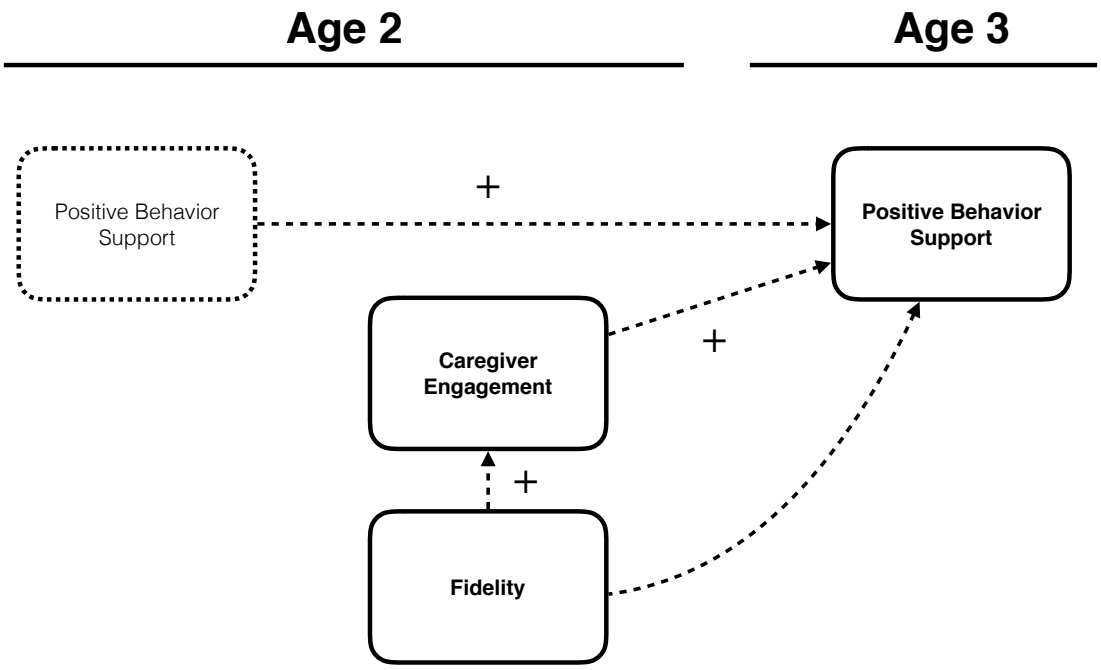

Figure 3. Conceptual model used in the stacked model to compare predictive validity of ratings based on audio review to those based on video review. The relation between fidelity, observed caregivers' engagement in the Family Check-Up (FCU) feedback session, and improvement in caregivers' positive behavior support one year later. Plus signs indicate paths identified as significant and positive by Smith and colleagues (2013) when tested as part of a more complex model; the direct effect of fidelity on positive behavior support was nonsignificant.

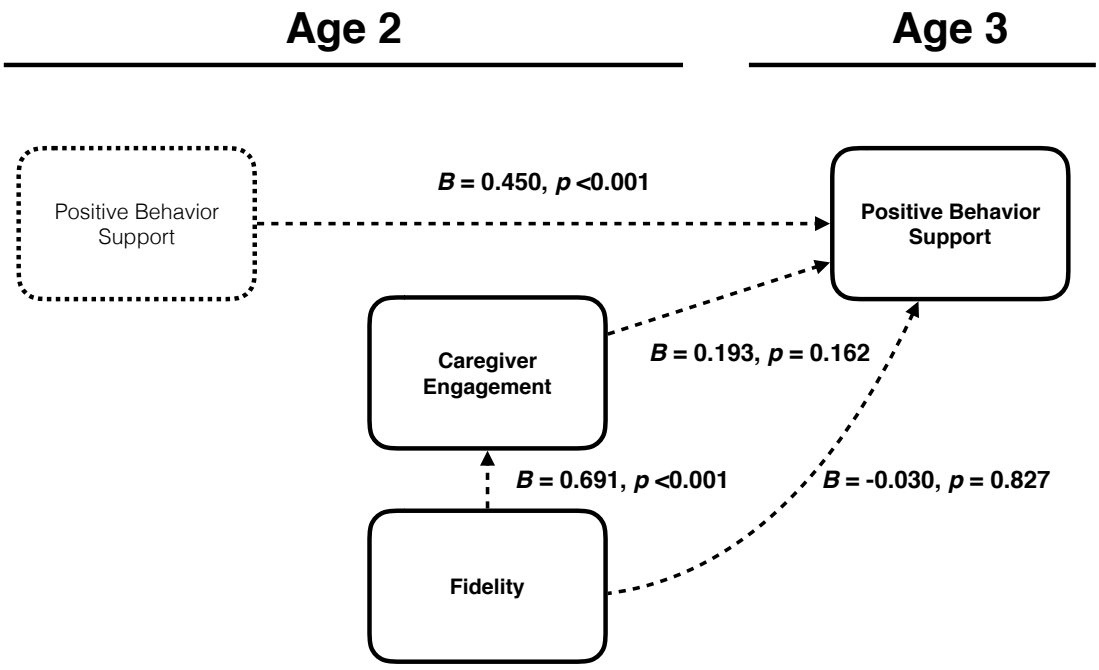

Figure 4. Standardized path coefficients and *p*-values for all paths in the model. Both fidelity and caregiver engagement ratings based on audio review and coded for the current study.

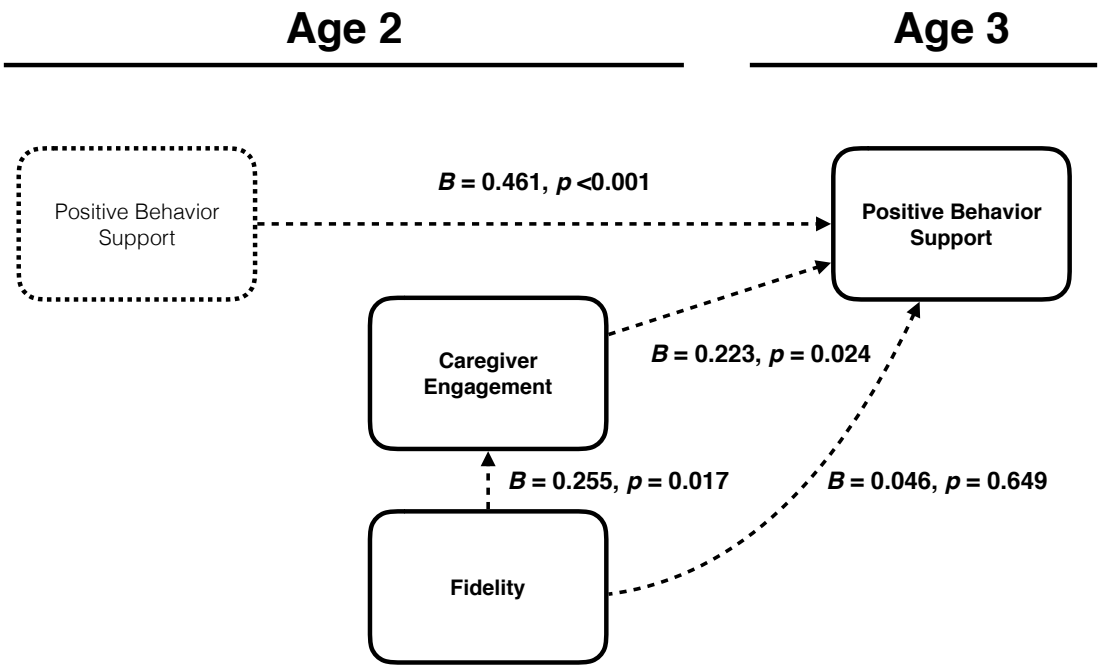

Figure 5. Standardized path coefficients and *p*-values for all paths in the model. Fidelity ratings based on audio review and coded for the current study. Caregiver engagement ratings based on video review and coded for Smith and colleagues' (2013) study.

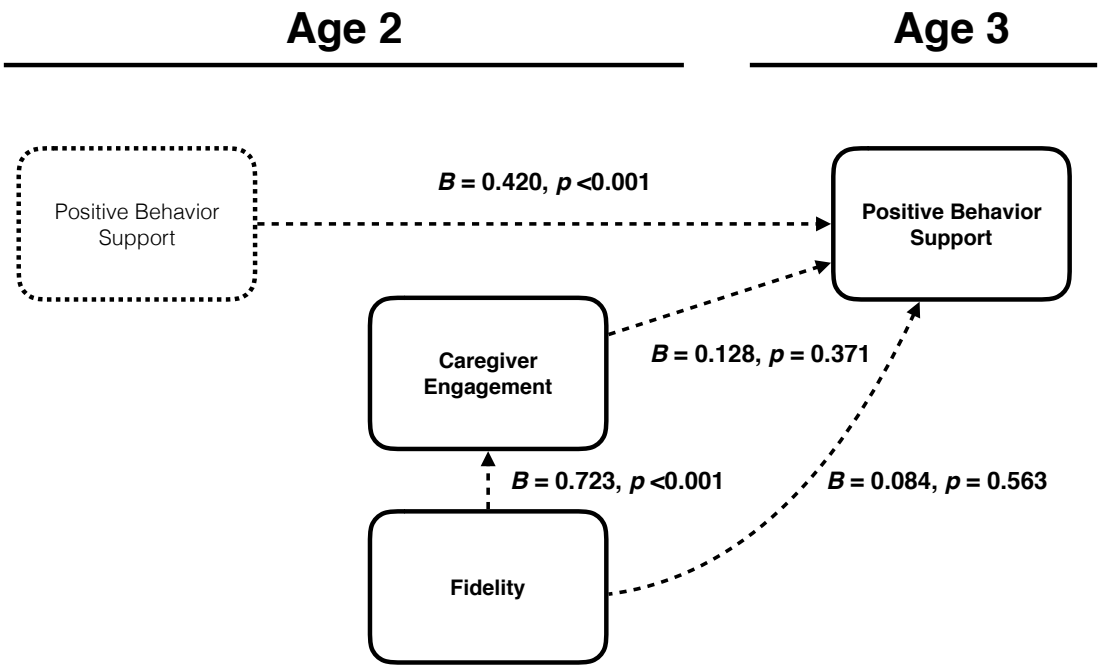

Figure 6. Standardized path coefficients and *p*-values for all paths in the model. Both fidelity and caregiver engagement ratings based on video review and coded for the current study.

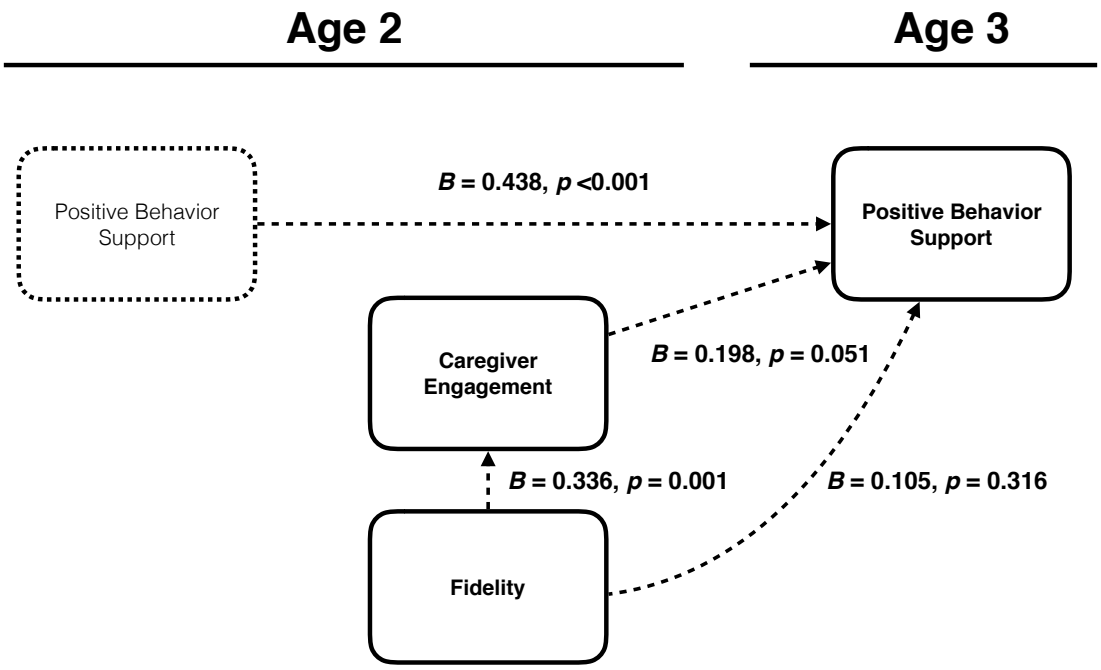

Figure 7. Standardized path coefficients and *p*-values for all paths in the model. Fidelity ratings based on video review and coded for the current study. Caregiver engagement ratings based on video review and coded for Smith and colleagues' (2013) study.

APPENDIX A

COACH MANUAL

Introduction

The overall intervention approach for which this fidelity system was designed is an ecological approach to family intervention and treatment (EcoFIT; Dishion & Stormshak, 2007). It consists of two linked intervention protocols, the **Family Check-Up** (FCU; Dishion & Kavanagh, 2003) and the family management intervention **Everyday Parenting** (EDP; Dishion, Stormshak, & Kavanagh, 2011).

The COACH is an observational system designed to assess the extent to which a provider displays adherence to the core components of the Family Check-Up and the Everyday Parenting program and implements these intervention models with skill. Research by leaders in the parenting intervention field has demonstrated that accurate implementation is relevant to high-quality services for families and is linked to benefits for the parent and the child (Forgatch, Patterson, & DeGarmo, 2005; Ogden, Forgatch, Askeland, Patterson, & Bullock, 2005).

The COACH Fidelity Rating System

The COACH fidelity rating system assesses clinical skills in five domains considered essential for the effective provision of the FCU and the EDP intervention sessions.

The COACH rating form is used to assess the extent to which the provider is:

Conceptually accurate and adherent to the intervention model

Observant and responsive to the family's needs

Active in structuring the session

Careful when teaching and providing feedback

Helpful in building hope and motivation

Using the COACH Fidelity Rating System

The COACH rating system includes two separate COACH forms for scoring the fidelity of intervention sessions:

1) The Feedback Session COACH Rating Form, designed for rating provider fidelity when delivering a Family Check-Up Feedback session
2) The Everyday Parenting COACH Rating Form, designed for rating provider fidelity when delivering parenting support interventions *following* the Feedback session using the Everyday Parenting program

Both versions of the COACH assess the same five broad domains, with the details tailored to best reflect factors unique to these two different phases of the broader EcoFIT intervention process. There is one important distinction between rating fidelity for a Feedback session and rating fidelity for an Everyday Parenting session. In order to rate provider fidelity in a Feedback session, the rater must first review the family's completed Feedback Form. This is because the Feedback session involves using the results of the child and family assessment, represented on the Feedback Form, to inform case conceptualization and recommendations for follow-up services. It is necessary to use information about the results of the FCU assessment and the case conceptualization when evaluating an intervention session. It is critical to review the assessment results summarized on the Family Check-Up Feedback Form (Dishion & Stormshak, 2007) before evaluating a session.

The domains of clinical practice assessed with the COACH rating system are described in detail below. In addition to measuring the five COACH domains when assessing provider fidelity to the Family Check-Up and Everyday Parenting models, it is also important to rate family engagement in the session. A key outcome of quality implementation of the Family Check-Up and Everyday Parenting sessions is that the family members present are engaged. Research has shown that ratings of providers on the five COACH domains are associated with ratings of parents' observed in-session engagement, and parents' engagement leads to positive intervention outcomes (Smith, Dishion, Shaw, & Wilson, 2013). Each COACH rating form includes a section for rating the client's level of engagement.

Each domain is scored on a scale of 1 to 9, with higher numbers reflecting higher levels of fidelity and clinical competence with the model. A score of 1, 2, or 3 indicates that the provider needs to work on improving in that domain. Scores indicating that a domain "needs work" means the

provider does not use the recommended process skills and does not display a clear understanding of the principles of the model. The provider will need to receive more training and consultation in order to implement the intervention with fidelity. A score of 4, 5, or 6 indicates that the provider is doing competent work in that domain. Scores in the “competent work” range indicate basic competence and adequate performance, which means that the provider has an acceptable level of skill and conceptual understanding but that this is accompanied by occasional errors or missed opportunities. A score of 7, 8, or 9 indicates that the provider is doing excellent work in that domain. Scores in the “excellent work” category indicate that the provider has mastery of the process skills that are required to deliver the model and has a clear understanding of the model’s principles and conceptual underpinnings.

The COACH rating form is divided into two columns. The first column highlights the criteria for effective practice for each domain. The second column indicates barriers to effective practice for each domain. When rating each domain, the rater should first consider to what extent the provider demonstrated the criteria for effective practice. Next, the rater should consider to what extent the provider’s delivery of the session demonstrated barriers to effective practice, which would lower the overall rating for the domain. As an example, if a provider generally demonstrates competence in the Careful When Teaching domain (and would have received a score of 7) but also provided a rationale that was unscientific, then the final score for Careful When Teaching might be 6. Engaging in many of the “barriers” would suggest a significant misunderstanding and perhaps the need for more direct instruction on the intervention model. When creating a domain score, the rater must consider each of the criteria for that domain.

The purpose of using the COACH rating system is to provide feedback and support to providers in the effective and accurate use of the FCU and EDP models. In general, providers that are <u>not</u> using either the FCU or EDP model should receive a score in the 0 to 3 range. That is to say, the COACH rating scores are based on positive features unique to the FCU and EDP models and are thus meant to reflect a provider’s skill in delivering these two intervention models in particular. While excellent clinical skills are necessary in order to deliver these models with competence, they are not sufficient; the provider must also demonstrate adherence to the unique features of the two intervention models. A provider who is struggling with the model (perhaps conducting their first Feedback session) might receive a score of 3, having demonstrated a few of the skills within each domain. We can expect that novice clinicians, or even experienced providers new to the model,

may not demonstrate sufficient skill to meet criteria, in which case their scores will be in the 3 to 4 range as a normal part of the learning process.

The COACH is primarily used as a tool for effective supervision, as it gives the supervisor specific criteria for guiding providers towards implementing with fidelity all phases of the EcoFIT model (i.e., both the FCU and EDP). The COACH rating protocol offers a venue for specifying provider areas of strength and areas for growth. This information can lead to improvement in performance. The level of detail provided by the COACH allows supervisors to effectively identify and therefore directly address areas where the provider does not understand the model, misses key opportunities to promote family wellbeing, or otherwise makes mistakes when implementing this model. It also helps supervisors to gain a clear understanding of the provider's strengths. Using the COACH in supervision gives providers useful feedback on how they can further develop their clinical skills, which will likely increase their confidence in using the intervention and promote their ability to implement with fidelity. Newly trained providers can use the COACH in supervision to support them in the learning process, and providers who are already certified in the model can use the COACH to help them maintain fidelity. Because high levels of intervention fidelity are associated with better outcomes, implementing with fidelity can be personally rewarding to providers because they see the positive effects of their work with families. Providers who implement with fidelity will observe improved family engagement in session as well as better attendance at future sessions. They will also observe increases in parent effectiveness and improvements in child behavior.

In order for supervisors and fidelity raters to effectively score an intervention session using the COACH rating system, the entire session must be reviewed.

Below, you will find a detailed description of each domain. Each domain is presented with an explanation of the key concept it represents and a list of the associated process skills. We then provide examples of actual clinical work that reflect exemplar (conceptually accurate) and nonexemplar (conceptually inaccurate) execution of each concept.

COACH DOMAINS: Descriptions and Exemplars

1. Conceptually accurate and adherent to the model

The provider demonstrates an accurate understanding of the model in terms of its emphasis on caregiver leadership in the change process and parent and child strengths, as well as support of the skills that define family management (i.e., positive behavior support, monitoring and limit setting, and relationship building). The provider follows the protocol for the session and refers to assessment results throughout. The provider tailors intervention in accordance with the results of the assessment. The provider addresses any factors that may interfere with successful engagement in family management practices and does not avoid focusing on these concerns just to steer clear of potential family-provider conflicts.

Recall that the EcoFIT model (i.e., the Family Check-Up followed by Everyday Parenting) is assessment-driven and intervention is tailored to the specific needs of children and families. This unique aspect of the model shapes all provider-parent interactions. For example, focusing on "parent self-care" may be conceptually accurate when assessment results suggest that a child's problem behavior is embedded in a parent-child dynamic strongly influenced by the parent's depression; however, if the child is showing problem behavior only at school, focusing on the parent's self-care may indicate a lower level of fidelity in this domain.

Key process skills that reflect conceptual accuracy and adherence to the model

- The provider follows model protocol and principles in the structure and content of a session. For example, it is important that the provider start the Feedback session by asking the parent what new insights about themselves, their child, or the family dynamic they might have gained from completing the Child and Family Assessment. (This is sometimes referred to as the "caregiver self-assessment".) The provider next explains the Feedback Form. The provider also gives data-based feedback using the Feedback Form and videotaped Family Interaction Tasks and ends the session with a discussion about goals and service options. At an EDP session, the provider presents a new skill by first offering a rationale and instruction, then modeling the skill, then having the parent practice the skill. The provider debriefs with the parent after each role play.
- The provider demonstrates competence in the use of FCU and EDP strategies and applies them appropriately to the family's situation in an assessment-based, tailored fashion. For example, when working with overly punitive parents, the provider will not begin with the

module on limit setting but rather will seek to first build the parents' competence in positive behavior support. Conversely, when working with parents who acknowledge and reward positive child behavior but are weak on setting limits, the provider provides feedback about this dynamic and begins follow-up services with a focus on limit setting (Dishion & Stormshak, 2007). At the Feedback session, the provider offers a menu of evidence-based services that address the family's specific needs.

- The provider treats the parent as expert, leads from behind, and uses motivational enhancement strategies to encourage caregiver engagement in the change process.
- A key component of being conceptually accurate in this model is that the provider remains committed to creating a shared perspective with parents about child problem behaviors and family challenges that are framed using a parenting or family management focus. For example, if concerns arise about a parent's depression, relevant links are made to other areas of child and family functioning, specifically parenting, as indicated by the assessment data. For instance, if observational assessment reveals that the parent is unable to adequately provide positive reinforcement to the child; the provider could explain that treating parent depression may naturally increase positive interactions with the child.
- The provider links feedback and treatment session goals to the assessment data and, during an EDP session, to the goals stated by the parent during the feedback session. The provider also demonstrates skill in making connections among assessment domains.
- When delivering the EDP, the provider describes key components of the EDP in ways that are meaningful to the client. The provider demonstrates accurate understanding of EDP model components and uses that knowledge to convey relevant examples and information to the client on the basis of the client's situation (e.g., single parenthood, poverty, recovery from addiction). Additionally, the provider should tailor how family management skills are presented and taught to ensure they are relevant for the child, taking into consideration, for example, the child's age and developmental level.
- The provider demonstrates understanding of behaviorally-oriented parenting interventions. This means the provider uses behavior-based rationales when discussing child behavior and when teaching parents new skills.
- When teaching a new skill, the provider offers a rationale and instruction, models the skills, has the parent practice using the skill in role play, and debriefs with the parent following each role play.

Barriers to effective practice

- The provider avoids providing feedback on areas that need attention or otherwise minimizes these challenge areas.
- The provider avoids direct discussion of the parenting practices identified in EDP
- The provider delves into tangents, engages in speculations or makes suggestions that are not evidence-based, or otherwise presents information or ideas that detract from the feedback process or EDP.
- The provider shows a premature focus on blaming or on a family's stories that detract from the feedback process or EDP.

Clinical Examples

Feedback Session Exemplars

- "Today is an opportunity for you to hear more about all of the information you shared with us on the questionnaires and the videotaped tasks so you can help Juan be more successful at home and in school."
- "Based on what you and Sean's teacher told us about his behavior, he is having more problems than most children focusing and paying attention in school. Does that fit with how you see things?"
- "Based on your report on Lucia's level of problem behavior, this is a very good time for you two to work on using positive behavior support and setting limits."
- "Based on the information you gave us in the assessment, using positive behavior support to reinforce Jamie's good behavior would be a good first step."

Everyday Parenting Exemplars

- "So, let's talk for a minute about what we call positive behavior support. We know from research that children learn positive behavior when their parents reinforce their good behavior. When you notice positive behavior, praise them or reward them with something positive, kids are more likely to repeat that behavior. Kids want your praise and encouragement and rewards. Giving these is a way you can support their good behavior."
- "It will be easier if we start by focusing on something Lucia is already doing some of the time. For instance, when she puts her shoes away when you ask her to, praise and

reward her. We can move to the other problem areas once she is following this kind of direction more consistently."

- "Last time we met we decided that this week we'd talk about using Quiet Time. Just as a reminder, Quiet Time is something you can set up to help your children behave better. It's a way to respond when they are misbehaving, because it interrupts their misbehavior."
- "You seem to feel that it's very important that you are the one to set the rules in your house, and at the same time, you worry about not taking your child's feelings and input into consideration. Is that right?"

Feedback Session Nonexemplars

- "So today we're going to go over the data from the assessment. I'll tell you about how Juanita is doing and then I will tell you about the next step in terms of treatment."
- "I see that Jonathon has been struggling at school. You mentioned that you've already gotten him tutoring services, so I don't think we really need to talk about this today."

Everyday Parenting Nonexemplars

- "A lot of parents struggle with Quiet Time, and, between you and me, I don't find that they work for me either – the trick I like is the 1-2-3 counting method. Would you like to try that?"
- "Sounds like praising his good behavior just isn't something you want to do, so let's move on to the next section – limit setting."

2. Observant and responsive to the family's needs

The provider pays attention to the family's concerns and contextual factors and responds accordingly in the session. The provider applies the principles and specific strategies of the FCU and EDP in a way that takes into consideration the family's unique situation. The provider keeps the work moving forward by integrating the broad goals and the immediate concerns of the family. For example, a parent may struggle with making changes in parenting when a disrupted marriage or depression is a factor. The provider should be responsive and sensitive to these concerns and balance them with the parent's stated parenting goals. There is a critical difference between simply reacting to a parent's needs without a plan versus observing and responding to the parent's needs with an explicit plan to focus on family management skills. On some occasions, being observant and responsive to a family's needs involves redefining the

focus and the work in the session to address the family's reactions to a previous session. Being responsive to the family's current problems and building a realistic foundation is an important competency for providers working in this model.

Key process skills in this domain

- The provider establishes a collaborative set, using reflective listening and empathy to build rapport and gain an accurate understanding of the family's perceptions of the problem and of their life context. At the highest level of competence, the provider offers observations of the parent's emotional expression (facial expression and/or posture), in addition to reflections of the content of what the parent is saying, to help the parent develop a better understanding of their underlying feelings and motivations. For example, imagine a mom and her fiancé (the future stepdad) at a feedback session, talking about getting married. If that future stepdad seems nervous or anxious, the provider can comment on his nervous affect, normalize it (e.g., "Getting married is a big step and it's only natural that you might be feeling a bit nervous."), and allow the couple to talk about their deeper feelings and the ways these relate to parenting the child who is the focus of the Family Check-Up.
- The provider tailors the session to the parent's level of education, emotional needs, life stress and contextual concerns, and cultural background.
- The provider uses the same language and examples the parent uses in order to reflect the family storyline and social context.
- The provider checks in periodically with the family to ensure that the session and intervention agenda fit the family's perceived needs. The provider incorporates the parent's immediate concerns and contextual factors into the session and adjusts the session agenda and intervention methods accordingly, when appropriate.

Barriers to effective practice

- The provider isn't responsive to the parent's input or behavior in session.
- The provider lectures or steam rolls; there is disproportionate therapist/client talk (goal is 1:1 ratio).
- The provider misses potential issues related to harm reduction or immediate action.

Clinical Examples

Feedback Session Exemplars

- “You told me that you’ve been feeling very down these past few months. Your scores on the questionnaire suggest that you are struggling with a depressed mood. It can be very difficult to parent when you’re feeling so low.”
- “You reported that the two of you are not on the same page when it comes to setting limits with Amanda. You also said that the two of you are fighting a lot lately. It can be very challenging to co-parent when you’re having trouble getting along as a couple.”
- “Linda, it’s been important for us to spend some time understanding your recent struggles with depression. I appreciate all that you shared, and we will come back to this important concern in a little bit. Also our purpose today is to help us both see how all of these areas connect for you and for your children, to help you and them get along better; so what do you think about us moving forward to this next area, looking at your parenting and Johnny’s behavior?”

Everyday Parenting Exemplars

- “I wonder if you might have some doubts about using Quiet Time. Can you talk a bit about why it works or does not work in your family?”
- “Your daily stress level is high. It makes sense that it would be very hard to give Andy clear directions when you are feeling stressed. Is that how you experience it?”
- The provider uses the word “we” to discuss setting an agenda. For example, “Last time we talked about making effective requests and we decided to have you track your requests. Can you fill me in on how that went? What did you notice about your requests and Juan’s behavior?”
- “It seems that both of you have many strengths that you bring to parenting, but your family is going through change. You’ve said that Ramon is upset about these changes and wants to have some input. I think a good place for you to start would be to use communication and problem-solving skills. This will give you the tools to solve some of the problems that come with blending two families, and it will also give Ramon and the other kids a chance to have some input.”
- “So it wasn’t possible for you to complete the home practice experiment from last week. I understand things come up. May I ask what got in the way? This way I can better understand your situation and we can make changes if we need to.”

Everyday Parenting Nonexemplars

- The provider interrupts the parent, cutting off what they say or completing their sentences for them.
- The provider, even with the best of intentions, lectures the parent about the parent's own experience.
    - "I know you've had a lot of difficulty in your life, and I think it's just wonderful that you're taking all these steps to make things easier for your children. I see how you've come a long way. You didn't have all this support from your parents, but now you're doing so much to support Emily. I can tell you've put a lot of thought and effort into this, and I'm sure it will pay off. I'm so proud of how far you've come. When we met, you didn't know where to begin, and now you're doing lots of fun things with your family. You were saying you just played kickball the other day, and I think those little things make a big difference. Things like that are so good for you …" (and so on).
- "I know you have been struggling with discipline, but it really shouldn't be that hard. Here are some handouts that you can read. This will make it easier for you to understand the steps. Once you use these steps, I think you'll find that Adam will stop leaving the house in the middle of the night."

3. Active in structuring the session

The provider actively structures the change process by using an assessment-driven case conceptualization and encourages parent involvement by using active strategies to teach family management skills. The Family Check-Up model requires the provider to be an active, yet flexible, leader in the session. In addition to listening and demonstrating support, and empathy, the provider can use actions such as posing provocative questions, conducting role plays, and redirecting discussions to motivate and empower parents to behave differently in their interactions with their children.

Key process skills in this domain

- The provider demonstrates good leadership and fidelity by following the structured steps of the Feedback Session and EDP intervention sessions.
    - The Feedback session begins with the parent self-assessment (i.e., the provider invites the parents to reflect on their experience during the assessment and offer new insights they may have had about their parenting or family management practices), then moves on to the provider explaining the Feedback Form and giving feedback,

and ends with a summary of key points, goal-setting, and discussion of potential follow-up services.
    - In Everyday Parenting, the provider structures sessions by presenting material from the program manual and using this material to teach parents new skills.
- The provider makes effective transitions from one activity or topic to the next, paces the session to maintain the family's attention and engagement without going too fast or slow, and weaves in suggestions for interventions or instruction as the opportunity to do so arises.
- Particularly for an Everyday Parenting session, the provider actively and collaboratively sets an agenda, having come to the session with a plan in mind based on the family's goals and previous session work. For a Feedback session, the provider has in mind a case conceptualization that guides the delivery of feedback particularly in regard to how connections between domains are drawn.
- The provider keeps the family "on track" or "on task" during a session by referring to the collaboratively established agenda as needed. If parents present distractions or side stories, the provider can redirect the focus back to the agenda or highlight a detail that relates back to the session goals.
- The provider comes to sessions prepared with the appropriate materials (e.g., video clips, handouts, worksheets) and uses them effectively.
- The provider asks for the parent's perspective and invites contributions and responses. When more than one parent or family member is present, the provider carefully balances support and attention among them and acknowledges that each participant has something of value to offer.

Barriers to effective practice

- The provider mismanages time; the session runs long or significant sections of the session are left unrealized (e.g., not all feedback domains are addressed).
- The session is disjointed and would benefit from better pacing, smoother transitions, and time for instruction.
- The session structure disrupts the client's potential to understand the feedback or EDP material and respond or engage.

Clinical Examples

Feedback Session Exemplars

- **Parent Self-Assessment: "**Through this process – completing questionnaires, doing the videotaped activities, and talking with me about your child – parents often feel that they learn something about their child or about themselves. What stood out for you?"
- **Explaining the Feedback Form: "**I'd like to show you this form that we will be using in our conversation today. This is the Family Check-Up Feedback Form, which is really like a map for where our conversation will be going today. This top part will reflect information you shared about your own wellbeing and the support you receive, the middle part focuses on Joni's behavior and wellbeing, and this bottom section covers different areas of family management. We'll be talking about the strengths you and Joni have as well as any concerns you mentioned in each of these areas. As you can see, the form is set up with colors, a little like a traffic light. Areas with marks in this green zone are strengths compared to other children or parents and their skills; those are things you want to keep going. When the mark is in the yellow zone, that signals a need for caution – we can pause and think about whether this is something to change in some way or just keep an eye on. An area with a mark in the red zone is an area that needs attention. This might be something to learn about or get help with. " The provider shows a blank form during the explanation and brings out the completed form when they begin to go over the assessment results.
- "I hear you both expressing a lot of concern about Vince's acting out, both in school and here in the neighborhood. From what you just told me, it sounds like he's gotten in with some boys who are good at getting into trouble. You're right to be concerned about this, and if you like, this is a topic that our follow-up sessions can address by helping you come up with strategies to keep tabs on Vince's whereabouts and to structure how he spends his time. How does that sound to you?"

Everyday Parenting Exemplars

- "At the Feedback session, we talked about focusing on Eli's homework routine at your house and working toward consistency across both households. Is this something you'd like to work on today?"
- "Last week you and I agreed that today we would work on limit setting. Before we start, I'd like to know if anything happened this week that we might need to consider while setting our agenda."

- "Danielle may not respond exactly how you want at the beginning, but with time and consistent encouragement, she will get into the new routine."
- "It seems like you are upset today about the fight you just had with the school. Should we go over that first so you can make a plan about how to deal with them next week or do you want to start where we left off on the home incentive plan?"
- "It sounds like there was a lot of conflict between you and Shandra's grandma this week. How did that affect your attempts to give Shandra more positive attention?"

Feedback Session Nonexemplars

- The provider does not ask self-assessment question, thereby missing opportunity to learn about parent perspective and establish value of parent voice in process.
- The provider begins Feedback by showing the parent their completed Feedback Form. A parent may have trouble listening to the explanation of the form and difficulty focusing on discussion of feedback in just one area if they can see the entire, completed form. Parents sometimes become so focused on looking over the marks on the form that they tune out what the provider is saying. It is best to use a sheet of paper to cover the part of the form that shows domains that have not yet been discussed.
- "As you can see, Vince is in the red zone here; that's very troubling. We know from a lot of research studies that kids in the red are headed for trouble, drug abuse, even jail. I know you don't want that for your kid, so it's time for you to take some action. Now let's look at how he's doing in terms of his relationships with other kids his age."

Everyday Parenting Nonexemplars

- The provider gets pulled off track into a family crisis or an interesting distraction. The provider loses the focus of the session and/or drops the agenda.
    - "I can't believe how your ex has been treating you! It seems unfair, and I can't imagine how his behavior is affecting the kids! Are you talking to them about this?"
    - "So it sounds like you've had a lot going on this past week. Say more about all that's been going on for you. We may not have time to get to the parenting material we talked about, but that's ok."
- The provider does not get the parent's input when setting the agenda for the session. For example, the provider might say, "Today we're going to talk about clear directions. The first thing to know about clear directions is …"

4. Careful when teaching and providing feedback

Because the Family Check-Up is a strength-based intervention, one element of careful teaching is identifying and building on existing parent, child, and family strengths. The provider offers feedback and guidance that is tailored to the parent and builds on their strengths. Careful teaching will increase the parent's accurate self-appraisals and motivate the parent to take corrective action and to build on existing strengths. The provider capitalizes on existing strengths to promote skill acquisition and development. The provider sensitively provides feedback and guidance to engage the parent in skill-building activities. The provider engages the family in an active learning process by using role play, by assigning home practice, and by facilitating the problem-solving process when the family encounters difficulty with implementing new skills.

Key process skills in this domain

- The provider notices, acknowledges, emphasizes, and builds upon existing family, parent, and child strengths.
- The provider gives clear explanations and demonstrations, both tailored to the parent's understanding of what is being taught. The provider breaks new skills into small, teachable units. Experiences of even small success build the foundation for increased confidence, motivation, and likelihood of future success. When a parent is having difficulty implementing a particular skill, the provider breaks it into smaller units.
- The provider draws a connection between the family assessment, previous sessions, and the current session. At a Feedback session, the provider refers back to what was said at the Initial Interview and what was reported and observed in the Assessment. At an Everyday Parenting session, the provider can refer back to these sessions and to previous Everyday Parenting sessions.
- The provider gives accurate, research-based rationales and uses evidence-based intervention procedures. (For examples of these, see the Summary of Research document on the Family Check-Up website.)
- The provider gives video feedback on parent-child interaction. It is best for the provider to show video clips of both successful use of a parenting skill and a missed opportunity. The provider gives a brief introduction for each. After playing a clip, the provider asks the parent what they noticed and then facilitates discussion about the relevant parenting skill.

- When delivering the EDP, the provider engages the parent in an active learning process. The provider uses role play to empower parents to learn a new skill by practicing it in session. The provider follows these steps when teaching a new skill:
    - Describes what the skill is and how it is used
    - Offers a rationale for using the skill
    - Models correct use of the skill
        - In addition to giving the parent an example of the right way to use the skill, the provider can also model common mistakes. The provider uses examples of the "wrong way" and "right way" to help parents discriminate effective from ineffective parenting. The provider may also show videotaped examples.
    - Gives step-by-step instruction that is tailored to the parent's learning ability, confidence, and willingness to participate
        - Establishes clear and realistic scenarios
        - Uses real situations generated by the parent
        - Breaks the role play into small steps
        - Gives clear behavioral prompts and directions (e.g., tells role play participants where to sit or stand, what to say, and how to say it)
    - Gives supportive coaching to the parent before and during the role play
        - Gives specific and encouraging feedback
        - Gives about four praise statements for every one correction
    - Debriefs after the role play, highlighting key points and validating parent effort and success
        - After practicing a skill the "wrong way" and the "right way", the provider asks the parent to compare how they feel when doing something the right way versus the wrong way and how their child might respond to these very different parent behaviors.
        - The provider asks questions. For example: How did it feel from the perspective of the parent? From the child's perspective? What did they notice? What did they like or not like? What parts were hard and what parts were easy?
        - The provider shares positive comments about what the parent did in the role play.
        - The provider offers praise before and after giving corrections.

- The provider discusses with the parent what barriers the parent might encounter when implementing a skill and how to address these potential problems. The problem may be child refusal, lack of support from the other parent, or incorrect application of the new skill. The provider normalizes common barriers and explains that making mistakes and encountering problems is a normal part of the change process.
- When delivering the EDP, the provider ends the session by assigning home practice. The provider assigns home practice that supports the work of the session and seems likely to be completed. The provider gives relevant handout materials and clear instructions about what the parent is expected to do between this session and the next. The provider asks the parent whether the parent expects to be able to complete this home practice successfully. If the parent is not confident about their ability to do the home practice, the provider offers encouragement and facilitates a problem-solving discussion.
- When delivering the EDP, the provider follows up on previous home practice assignments, demonstrating interest in the parent's progress. When home practice assignments have not been completed, the provider and parent work together to identify and address the barriers to success.

Barriers to effective practice

- The provider minimizes or avoids areas of concern or opportunities for feedback and correction.
- If an EDP session, the provider focuses on an EDP module that is not appropriate for the family (i.e., does not meet the needs identified in the FCU Assessment).
- The provider gives rationales or suggestions that are unscientific, inaccurate, or unprofessional.
- The provider teaches too much in the session (i.e., lectures or gives too much information).

Clinical Examples

Feedback Session Exemplars

- "Let's take a look at the two of you playing that game together. I'd like you to watch for Vince's good behavior and your own positive parenting here."
- **"**You told me that one of your best parenting skills is noticing his positive behavior – this is such an important skill because it motivates children to stay on track with good behavior. You've also said you'd like to be more active in structuring his time after

school. I think this is a great next step. You already know how to praise his good behavior. Let's talk about how you can use that skill to get him participating in more structured activities."

- "When you structure his activities and provide more monitoring of him, it makes it less likely that he will have time to get into trouble with the neighborhood boys. We also know that kids who are well-supervised tend to do better in school both with their behavior and their grades."

Everyday Parenting Exemplars

- "You made eye contact and used a calm tone of voice. Now let's try to make the request more specific. What exactly do you want him to do? What are the steps?"
- "Let's start by looking at the common pitfalls most of us fall into when giving directions. For example, how many parents do you think shout a command to their kids from the other end of the house? Let's try that, just to see what it feels like."
- "Let's try something different now. Don't worry about getting it right the first time. I'll be here to help you out."
- "Let's take a look at this videotape on positive behavior support, and then we'll talk about how you think it might work for you and Lucia."

Feedback Session Nonexemplars

- The provider does not refer to the family's strengths or mentions strengths briefly without discussing them to the same extent as the provider discusses problem areas. The provider spends more time and uses stronger language discussing problem areas.
- "Setting limits is a really important thing to do"
- The provider does not give video-based feedback (i.e., does not show a clip or does not use the clip to generate discussion).
- The provider avoids giving feedback on challenging areas or dismisses or minimizes concerns. For example, the provider might say, "Well, you and his teacher both said that he can be aggressive, but I didn't see that on the videotaped tasks and you just said that he's getting better. Sometimes kids just outgrow these things."

Everyday Parenting Nonexemplars

- The provider emphasizes problems over strengths. For example, the provider might say, “I think the main problem is just being consistent. In the office last week, you were able to make a request and give a reward, but it just isn’t happening at home.”
- The provider lectures and/or provides more information than the parent is ready to take in.
- The provider provides too little information, leaving the parent with a poor understanding of the skill or topic area. For example, the provider might say, “Oh, so you’ve done behavior charts before? That’s great! So you know what you’re doing! Go ahead and put Allison on a chart for her behavior, and then next time we can talk about how it worked.”

5. Helpful in building hope and motivation

The provider promotes hope, motivation, and change by using motivational enhancement strategies. The provider does the following: (1) gives feedback to the parents, (2) emphasizes that the parent is responsible for the change process, (3) teaches the parent effective change strategies, (4) gives the parent a menu of service options and does not control the change process by offering only one option, (5) expresses empathy for the family’s situation, and (6) promotes the parent’s self-efficacy. These process skills are used in moment-by-moment interaction with the family. The provider uses these skills to build hope that the intervention will be successful. By using these skills, the provider invites the parent to be an agent of positive family change and motivates the parent to work toward this positive change. The provider frames the family situation in a way that highlights the family’s strengths. The provider asks questions that help the parent reflect on their behaviors and goals. The provider offers supportive statements that validate the complexity of the change process in families.

Key process skills in this domain

- The provider uses a number of clinical process skills to build hope and motivation. The provider offers validation, empathy, and hopeful reframing. The provider employs humor. The provider paraphrases and summarizes what the parent says, appropriately normalizes the parent’s concerns, and asks questions to explore the parent’s motivation to change.
- The provider elicits change talk (i.e., the parent’s acknowledgment that change is possible) by highlighting discrepancies between the parent’s values and behavior and between the parent’s hopes and concerns, exploring both sides of the parent’s ambivalence.

- The provider identifies the parent's strengths and invites the family to reflect on previous successes in various domains.
- The provider supports the parent's self-efficacy by helping the parent to identify realistic goals with achievable steps, by assigning home practice with achievable steps, and by expressing optimism about the parent's ability to make the desired change.

Barriers to effective practice

- The provider misses opportunities to highlight past efforts, successes, or strengths.
- The provider gives advice, shows disagreement, or teaches in the face of client ambivalence or discord.
- The provider undermines the parent's feeling of self-efficacy and hope, either through words or actions.
- The provider gives a discouraging message that undermines change (e.g., assigns blame, makes a critical comment, etc.).

Clinical Examples

Feedback Session Exemplars

- "You have had some experience making big changes in your life; what helped you to make those changes? What can you pull from those experiences to help you be successful now?"
- "You mentioned that it's really important for you that your son has a better chance in life, that you want him to graduate from high school and have the chance to go to college. You said you want to do what you can to help your son have a good life."
- "You've said you want to make sure Alicia follows the rules, but you're also concerned that she won't talk with you anymore if you enforce the rules by giving her consequences when she breaks them. What would have to happen for her to follow the rules and continue to go to you for support?"

Everyday Parenting Exemplars

- "It's important to remember that change doesn't happen overnight. I know it can be discouraging when you try and things aren't changing. Let's see if we can put our heads together and see what we might do differently."

- "You said that the time-out strategy didn't work for you. Let's take some time to explore how that went. What didn't work? What did work? What got in the way?"
- "You're right, setting limits can be hard. The fact that you are meeting with me and practicing these skills shows that you are motivated, and I can see you're making progress."
- "Last week, you said that you wanted to use Quiet Time as a consequence and not yell at Suzie so much. It sounds like you did try using Quiet Time once, but that it was hard for you, and the rest of the week was more yelling. What didn't work? … What did work? … What got in the way?"
- "What could we come up with that would remind you to notice his good behavior? Let's brainstorm some ideas."

Feedback Session and Everyday Parenting Nonexemplars

- The provider assumes client motivation: "If you're like me, you'll enjoy organizing your children's chores on this chart!"
- The provider moves into teaching too quickly, without taking a collaborative approach, such that the client is motivated to learn the new skill.
- The provider goes along with the parent's unrealistic goals and does not help the parent refine the goals to make them measurable and achievable.
- The provider ignores the parent's lack of success and does not explore barriers and help the parent to problem-solve.
- The provider undermines hope; for example, the provider responds to a mother's report of depression by saying, "And what really matters about that is that your depression can be affecting Sandi. The research tells us that when mom is depressed, children tend to have more behavior problems in school and at home."
- "Once you start being more consistent about the house rules, you will see an immediate change in your child's behavior."
- The provider places responsibility for lack of progress with a skill on the parent rather than offering support. For example, the provider might say, "It seems hard for you to complete the home practice activities. Can you tell me why?"

Parent In-Session Engagement

A key outcome of quality implementation of the Family Check-Up and Everyday Parenting is that the parent is engaged in the session. While fidelity generally increases engagement, this isn't always the case. For a variety of reasons, a parent may not be ready or able to engage fully in the change process. For example, life events, stressors or a parent's mood may make it challenging for a parent to engage in the session. A provider may implement the session with fidelity and still experience difficulty engaging the parent. Parent engagement may be a function of life events, developmental status, interpersonal style, mood disorders, and/or frustration with a personal situation. A provider can competently use many of the key process skills and still have difficulty providing feedback, teaching various skills, or connecting with the parent. The following checklist provides some guidelines for gauging the parent's level of in-session engagement.

High Engagement (scores 7–9)

Client actively participates in session by:

- Engaging in conversation with the provider and staying on topic
- Giving complete, thoughtful responses to provider's open-ended questions
- Offering verbal or nonverbal appreciation to provider
- Discussing a past success or expressing hope that a technique or idea will work
- Showing willingness to discuss personal information and problems
- Articulating problems, goals, or what they would like to do differently
- Actively participating in role plays
- Reporting on a completed home practice
- Bringing materials to the session, such as completed charts or report cards
- Asking questions and elaborating on the provider's observations
- Demonstrating an understanding of the provider's comments
- Displaying emotion related to topic
- Showing caring and concern for child and family
- Using change language
- Offering solutions to problems
- Taking responsibility for their role in family management
- Showing initiative
- Showing a willingness to try the provider's suggestions

- Being honest and genuine in working with the provider

Moderate Engagement (scores 4–6)

Client participates in session by:

- Showing modest or occasional verbal or nonverbal signs of engagement, such as head nodding, eye contact, sitting with an open posture, facing provider, smiling or using other expressions of affect (contingent upon cultural differences)
- Expressing some verbal or nonverbal agreement and elaborating on the provider's comments
- Giving short responses to the provider's open-ended questions
- Displaying some emotion related to a topic
- Showing some willingness to disclose personal information, discuss problems
- Expressing some ambivalence or reluctance about change

Low Engagement (scores 1–3)

Client appears inattentive or disengaged in session by:

- Appearing to shut out the provider by not paying attention to what the provider is saying or doing
- Not making eye contact, not smiling, or otherwise not expressing engagement nonverbally
- Displaying flat affect (contingent upon cultural differences)
- Displaying signs of boredom, such as looking around room or yawning
- Giving very brief responses to provider's open-ended questions
- Indicating chaotic or disorganized commitment, making it difficult or impossible to work systematically (e.g., responding to a cell phone during a session, getting into an extended interchange with an outside person, cutting the session short, bringing an unannounced person to a session)
- Showing reluctance to discuss personal information or problems
- Repeatedly side tracking to off-topic issues
- Providing few opportunities to engage in the session agenda
- Seeming to be dishonest or disingenuous while working with the provider
- Derailing conversation, such as repeatedly going off topic, going on tangents, interrupting, or talking over the provider

- Resisting the provider's presence or the work, such as being dismissive, denying any problems, expressing doubt about techniques or therapy in general
- Showing lack of ability to process provider's comments
- Showing signs of possible intoxication
- Revealing possible cognitive or learning difficulties that would keep them from engaging fully in the session

Observation Procedures and Training COACH Raters for Research Purposes

COACH raters (a.k.a., coders) are required to be well-trained and grounded in the FCU model and in the *Everyday Parenting* program (EDP). They should be familiar with the following intervention materials: *Intervening in Children's Lives: An Ecological, Family-Centered Approach to Mental Health Care* (Dishion & Stormshak, 2007) and *Everyday Parenting: A Professional's Guide to Building Family Management Skills* (Dishion, Stormshak, & Kavanagh, 2011). Having familiarized themselves with the concepts, coders will be able to recognize competent adherence to and delivery of the FCU and EDP sessions and evaluate the following: session structure, feedback and teaching strategies, use of role play, application of appropriate process skills, management of client resistance, motivational interviewing, and overall quality of the session. Feedback sessions require a structure that defines aspects of the provider-client conversation and how assessment information is used to tailor follow-up treatment, if follow-up treatment is recommended.

In general, the best ratings are achieved when the entire session is watched. This is especially true for the Feedback session. However, viewing times for the EDP sessions can be shortened by using the 15-minute segment that the provider deems the best example of his or her competence during that session. Coding may take up to 2 times the duration of the session or session segment. If the entire session will not be rated, the coder should rewind the tape to approximately 1 minute before the start time of the segment to provide some context for the segment; however, actual rating begins at the indicated start time.

COACH training takes approximately 20 hours before the coder is considered competent and reliable. Coders should be advanced undergraduate students in the behavioral sciences or