

Examining the Active Ingredients of Consultation to Improve Implementation of a Parent-mediated Intervention for Children With Autism in the Community Mental Health System

NCT04654117

06/17/2024

This study will examine which components of consultation are most helpful in improving clinical services that community providers deliver to Medicaid-enrolled children with autism spectrum disorder (ASD). Specifically, the project will focus on parent-mediated intervention, which is a best practice for improving skills in children with autism spectrum disorder. Previous work from our research team demonstrated very low frequency of billing claims for parent-mediated intervention within the Medicaid Autism Benefit – nearly half of children had not received a parent-mediated intervention session over the course of 6 months, and only 2.7% received parent-mediated intervention at a frequency that is consistent with evidence-based models (Straiton et al., 2021a). Qualitative data illustrated that providers are also largely unaware of evidence-based components of parent-mediated intervention sessions (e.g., care-giver practice with feedback). Our work highlighted that providers in this system are underprepared to deliver parent-mediated interventions and perceive numerous barriers at the family-, provider-, and agency-levels (Straiton et al., 2021b), such as perceived limited engagement and interest from caregivers, limited provider training, and competing demands on providers' time. This research demonstrated that providers desired training in the delivery of a manualized parent-mediated intervention that could be delivered via telehealth during the pandemic to address social communication deficits in children with ASD (Project ImPACT). The current project will address this need by supporting providers in this same system to use a manualized parent-mediated intervention with their clients with ASD. We will also be able to examine which components of consultation are most helpful in supporting providers to use this best practice within this system.

Specific Aim 1: Identify the potential active ingredients of the consultation model by evaluating its effects on providers' treatment adherence and parent-mediated intervention competence. I

predict that the feedback component will improve adherence and competence over and above improvements from the case support and skill rehearsal components.

Specific Aim 2: Examine the relative feasibility, acceptability, and appropriateness of each consultation component using a single-case design: component analysis. Feasibility is the extent to which a practice can be successfully carried out within a setting. Acceptability is the extent to which a practice is agreeable and satisfactory. Appropriateness is the perceived fit or relevance of a practice to address a problem. I predict that providers will perceive the case support component to be the most feasible, acceptable, and appropriate of all components.

Specific Aim 3: Examine the effects of the consultation model on case penetration and the feasibility, acceptability, and appropriateness of the intervention (Project ImPACT). I predict case penetration (i.e., total number of Project ImPACT cases on a provider's caseload divided by the total number of eligible clients) and intervention feasibility, acceptability, and appropriateness to increase over time.

Exploratory Aim 4: Demonstrate associated social communication outcomes for Medicaid-enrolled children with ASD from baseline to post-consultation. Given that consultation leads to improved adherence and child outcomes and Project ImPACT results in improved child social communication outcomes, I predict that our consultation model will be associated with improvements in child social communication skills.

Experimental Design and Procedure

This study sought to examine the active ingredients of group consultation using a component analysis design, which is a type of single-case experimental design that allows for causal interpretation of the effects of treatment elements within a treatment package. This study was approved by Michigan State University's Institutional Review Board. Methodology is largely consistent with published standards for single-case design (Kratochwill et al., 2013). In this study, we capitalized on the use of advanced statistics (i.e., multilevel modeling) to quantify the effects of consultation components on clinician fidelity within our ABCD component analysis with multiple baselines. Descriptions of the study design are provided below.

Design overview. Groups of 3-5 clinicians were recruited by each agency leader. Once all clinicians at an agency were enrolled, they had a 2-week initial training period in which they completed the Project ImPACT Beginner e-Course, a 6-hour self-directed online tutorial on Project ImPACT intervention content. The Beginner e-Course consisted of didactic modules with slides and audio recordings about Project ImPACT NDBI intervention strategies and best practices for collaborating with parents. Modules included video examples of caregivers implementing the intervention strategies and periodic quizzes about course content, but did not include explicit instruction about parent coaching.

Following the initial training period, clinicians entered the Baseline Phase (3 to 9 weeks; randomized by agency) during which they were asked to follow the Project ImPACT coaching manual and meet weekly with their enrolled families, but did not receive consultation support. After the baseline phase, all enrolled clinicians at each agency entered the consultation phase, during which they received 12 group consultation sessions, once per week, for 1.5 hours per session. The consultation phase included three, 4-week blocks in which the group completed one consultation training activity per block (i.e., case support, skill rehearsal, or feedback on

videotaped sessions). Clinicians submitted recordings of all Project ImPACT telehealth sessions throughout the Baseline and Consultation phases, and then again at Follow Up (8-weeks post consultation). Information about the procedures used in each phase is outlined below.

Baseline phase. Clinicians received an e-book copy of the Project ImPACT Coaching Manual, which includes detailed descriptions about the theoretical underpinnings of the intervention and lesson plans for each of the 24 sessions (e.g., session agendas, handouts, access to video models of NDBI intervention strategies, examples of trouble-shooting tips for common challenges with clients). Clinicians also received an e-book copy of the Project ImPACT Parent Manual, which includes descriptions of the NDBI strategies in parent-friendly language and tips about how to apply NDBI strategies at home and in the community. Clinicians were instructed to follow the lesson plans in the Project ImPACT coach manual and no consultation support was provided during the Baseline phase.

Baseline lengths were randomized by a statistician unaware of agency characteristics using a random-number generator. Published standards for single-case designs require a minimum of 3 observations per phase to “meet standards with reservations” (Kratochwill et al., 2013). See Table 2 for the baseline lengths for each agency. Baseline lengths for agencies 1-4 were randomized for 3-6 weeks. Due to a methodological challenge (outlined below), baseline lengths for agencies #5 and #6 were randomized for 6-9 weeks to ensure that clinicians had sufficient baselines.

Consultation phase. Group consultation sessions took place via a teleconference platform (Zoom). During a given consultation condition (i.e., 4-week block in which only one consultation training activity was delivered), no components of any other conditions were provided. For example, during the skill rehearsal condition, no feedback was given about the role

plays, so that the only performance feedback clinicians received occurred in the feedback condition. All providers received each of the 3 consultation conditions (i.e., training activities), with each training activity delivered in isolation for a given 4-week consultation period. The order in which providers received each component was randomized for each agency by a statistician unaware of agency characteristics using a random-number generator, such that each agency had an equal chance of receiving one of the predetermined permutations of conditions (e.g., equally as likely to be randomized to ABCD order as ADCB order). Consultation sessions were rescheduled for the following week if more than half of participants at an agency would not be present or if the consultant would not be present. See Table 2 for the consultation component order for each agency.

Case support condition. With one week's notice, clinicians were asked to brainstorm at least one implementation challenge that they experienced recently when delivering Project ImPACT (e.g., getting the caregiver to complete live practice during sessions, uncertainty about how to complete Practice Plan worksheets). At the beginning of each case support consultation session, the consultant screenshared a blank note-taking table and asked each clinician to list recent challenges they've had when implementing Project ImPACT. She asked all other clinicians to endorse whether they also experienced that challenge. Once all challenges were listed, she asked the group to choose 2-3 challenges to problem-solve during the present session. The consultant then asked the group of clinicians to brainstorm possible solutions for each selected challenge, noting their responses and any specific resources/handouts that were mentioned. Once the clinicians finished listing possible solutions, the consultant added her own ideas of possible solutions. Topics included challenges with telehealth, confusion about intervention content and/or handouts, and addressing client/family needs (e.g., difficulty

promoting caregiver engagement in session). Problem-solving specifically addressed ways to improve implementation of Project ImPACT within that particular agency's setting (e.g., within the resource constraints of that agency). Once the group agreed that they had no additional ideas for possible solutions, the consultant moved on to the next selected implementation challenge. Any challenges that remained were kept in the note-taking table to be discussed at the next case support session, if clinicians desired. Following the consultation session, the consultant emailed the note document to the clinicians.

Skill rehearsal condition. With one week's notice, the consultant sent an email that assigned a series of role plays that corresponded to the structure of a Project ImPACT session. The assignments included the following role play scenes: Scene A: setting the session agenda, reviewing the previous session's Practice Plan (homework sheet), and introducing the new NDBI technique (e.g., Prompts for Using Communication); Scene B: introducing the video model to demonstrate the new technique, screen sharing the video model, and helping the parent to reflect on the video model; Scene C: live caregiver practice with coaching comments (i.e., a videotaped parent-child interaction was screenshared by the consultant, and the consultant paused throughout to solicit coaching comments from the clinician); and Scene D: reflecting on the parent practice section and developing a new Practice Plan (i.e., homework sheet) for between-session practice on the technique. The email also included information about which session number the group would practice (e.g., Session 14: Prompts for Using Communication), background information about the pseudo client and caregiver (e.g., child's age, language level, and intervention goals), and the necessary paperwork for the session (e.g., completed Practice Plan for homework review section of the session).

The email also outlined directions for how each clinician would take turns pretending that they were the coach (picking up where the last scene ended), with the consultant role playing as the caregiver. The consultant reminded the clinicians to refrain from providing oral feedback and to instead write any feedback comments on a worksheet provided at each consultation session. These reminders were to ensure that the research design could clearly distinguish the effects of *rehearsing* parent coaching skills used in Project ImPACT sessions (i.e., Skill Rehearsal condition) from *receiving performance feedback* (i.e., Feedback condition).

During the Skill Rehearsal consultation sessions, the consultant began by reading the information provided in the most recent email and reminding the group that they should refrain from providing oral feedback to each other. She also explained that she would refrain from providing oral feedback, unless there was an important question about technology or program materials that needed to be answered in order to complete the role plays. Once a role play scene was completed, the consultant instructed participants to record their positive and corrective comments via written comments on a worksheet. She then asked the next participant to begin the following scene. This process was repeated until all scenes were finished. At the end of the session, clinicians were asked to email their completed worksheets to the consultant and not to share feedback with each other.

Feedback condition. With one week's notice, clinicians were asked to submit timestamps for a 5-minute clip from their most recent recorded Project ImPACT telehealth session. They were asked to submit the timestamped clip by 5:00 pm on the evening before the Feedback consultation session. The consultant asked clinicians to choose a clip in which they had a challenge delivering the intervention. To provide some structure about ideas for selecting clips, each week, the consultant assigned clinicians to submit specific sections of Project ImPACT

sessions (e.g., caregiver practice with coach feedback) if they did not have another moment in mind. These were the same sections/scenes that were used in the Skill Rehearsal consultation sessions, as outlined above. The consultant explained that clinicians were welcome to choose their own clip outside of those suggested assignments if they had a different moment that they would like to share. If a clinician did not provide timestamps by 5:00 pm on the evening prior to the consultation session, the consultant selected a clip for the clinician. Clinicians hardly ever provided their own timestamps; therefore, the consultant selected nearly every video clip and used this assignment structure to maintain consistency in how clips were selected.

During the group consultation session, the consultant asked the clinician to provide any relevant background information about the family or session prior to screensharing the clip. The consultant then screenshared the 5-minute clip and led the group to provide two rounds of feedback. First, the consultant asked each clinician to share positive feedback and/or specific praise about what they observed. Then, the consultant provided her own specific praise about the clip. Next, the consultant directed each clinician to give 1-2 suggestions of constructive feedback about the video clip. The consultant then provided her own constructive feedback comments. After both rounds of feedback were completed, the consultant asked the next clinician to introduce their video clip. She repeated this procedure until all video clips were shown.

Follow-up phase. After the consultation phase, clinicians spent 8 weeks in a follow-up phase in which they were asked to continue delivering Project ImPACT weekly with their client(s). Clinicians did not submit recordings of Project ImPACT telehealth sessions during this period, but were asked to submit one final recorded session and one final questionnaire at the end of the follow-up period. We received 7 sessions in the follow up period; 3 sessions happened to be Session 23: Update Your Child's Goals, which was atypical because it was a goal-setting

session and could not be used for parent coaching fidelity. We decided not to include the 4 remaining follow-up in analyses because there were so few of them.

Methodological challenge with the baseline phase. The research team became aware of a methodological flaw following visual analysis of manual adherence data from the first 4 agencies. Within-condition visual analysis indicated that baselines across agencies were unstable due to inflation of the manual adherence scores for the first 2 sessions of Project ImPACT, as compared to sessions 3-24. Conceptually, this made sense because these 2 sessions were atypical and did not include instruction on NDBI techniques nor an active parent coaching component in which clinicians provided feedback; session 1 was an overview of the program and session 2 was a goal setting session. Both sessions are much less complex than the other parent coaching sessions. Agencies #1-4 had started data collection by this point, and due to the initial randomization procedure using parameters of 3-6 weeks of baseline, only 1 agency had at least 3 weeks of usable baseline sessions (Agency #2). For this reason, an additional 2 agencies were recruited (Agencies #5 and #6), and baseline lengths for those agencies were randomized for 6-9 weeks to ensure that baseline lengths had at least 3 usable Project ImPACT sessions. Additionally, to allow for the retention of all clinicians in Agencies #5 and #6, baselines for those two agencies were extended by 1 week to allow for clinicians who had multiple cancellations in the baseline period to acquire at least 2 usable baseline sessions.

Consultant Fidelity to Consultation Conditions

Consultant fidelity to the three consultation conditions was assessed by 2 coders naïve to condition and timepoint using the Project ImPACT Consultation Adherence Checklist. The checklist included scores for the percentage of required training activities for each of the 3 consultation components, as well as the length of time spent on the training activities (e.g., time

spent in role plays). This measure has not been empirically tested, but was used to understand the level of experimental control in the study (i.e., extent to which only one consultation component was delivered in a given phase). Items were rated as Observed (1), Not Observed (0), or Not Applicable. Example items for the Feedback condition include: consultant asked each trainee to introduce their clip prior to screensharing it, consultant prompted each trainee to provide specific praise/constructive feedback about each clip shown, consultant provided specific praise/constructive feedback after each clip. Example items for the Skill Rehearsal condition include: each trainee present at the session completed at least 1 role play scene, consultant asked trainees to self-reflect on clinician performance in role play scenes using the reflection worksheet, scene A was completed. Example items for the Case Support condition include: consultant asked the group of trainees to provide potential solutions to the implementation challenge, consultant provided additional potential solutions to the implementation challenge. Coders also noted timestamps for various activities to determine the amount of time spent on activities of note (e.g., amount of time providing performance feedback).

Five agencies completed all 12 consultation sessions, while one agency completed only 6/12 sessions before all clinicians withdrew from the study. One video recording was corrupted and was unable to be scored; thus, there was a total of 65 consultation sessions that could be coded. Coders needed to meet training criteria of 3 videos in a row with 80% of all items in exact agreement with the master coder. Monthly meetings with the coding team were held in which all members independently coded a video, and the group compared their scores to the master codes and discussed discrepancies. Coders were randomly assigned consultation sessions to score. Ten of the 65 sessions (15%) were rated by both raters to calculate inter-rater reliability. Intraclass correlations were as follows: Feedback = 0.99, Skill Rehearsal = 1.00, Case Support = 1.00.

Case Support. Case Support sessions were implemented with 100% fidelity. Case Support sessions had an average of 38:53 minutes spent on case discussion and problem-solving implementation challenges. No Case Support sessions included time spent receiving performance feedback. Of the 20 Case Support sessions that were coded, 3 (15%) included feedback from the consultant about technology support such as screensharing procedures (mean = 1:30 minutes, range: 0 to 14:48 minutes) and 1 session (5%) included feedback from the consultant about the use of program materials (3:16 minutes).

Skill Rehearsal. On average, Skill Rehearsal sessions were implemented with 93% fidelity. Skill Rehearsal sessions had an average of 34:52 minutes spent in active role play. Of the 22 Skill Rehearsal sessions that were coded, 15 (68%) included some amount of feedback on role plays provided either by trainees or the consultant. Of note, the amount of time spent on *performance feedback* (i.e., feedback about how clinicians implemented the intervention during the role plays) was quite low; on average, the consultant provided 11 seconds of performance feedback (range: 0 seconds to 1:26 minutes) and trainees provided no performance feedback to each other. The majority of the feedback provided in the Skill Rehearsal sessions was related to the use of program materials needed for the role plays (mean = 23 seconds, range: 0 seconds to 1:43 minutes) or technology support regarding role play activities such as screensharing a document (mean = 1:31 minutes, range: 0 seconds to 6:26 minutes). Three of the Skill Rehearsal sessions (14%) included some Case Support activities, with an average of 29 seconds being spent on clinical suggestions about how to address a challenge with a case (range: 0 seconds to 10:05 minutes). Case support was only provided by the consultant if a clinician directly requested support and the consultant felt that refraining from responding to the request would diminish the consultee-consultant alliance and possibly affect clinician participation in the research project.

Feedback. On average, Feedback sessions were implemented with 97% fidelity, with no time spent on Skill Rehearsal or Case Support activities. Consultation sessions included an average of 18:09 minutes spent watching Project ImPACT session videotapes (range: 0 seconds to 26:06 minutes), 38:20 minutes spent on performance feedback (range: 0 seconds to 48:31 minutes), 1:57 minutes spent on feedback about technology support (range: 0 seconds to 7:23 minutes), and 2:08 minutes spent on feedback about the use of program materials (range: 0 seconds to 8:13 minutes).

Data Collection for Observational Measures of Clinician Fidelity

Clinicians recorded all telehealth sessions with enrolled families and submitted the video recordings via a HIPAA-compliant Dropbox link. There was a high frequency of session cancellations across the study, which is common in the community mental health system. For the 3 agencies with sufficient baseline lengths, 76 sessions of the 299 scheduled sessions were canceled (25.4%).

Fidelity. Adherence and competency were measured independently, as research suggests that providers may adhere to manual content without demonstrating appropriate competence in the delivery of the content (Cross & West, 2011). For instance, a less competent provider could give constructive feedback to a caregiver, but might do so in a way that is too harsh, making the comments unhelpful. In that case, they might be adhering to procedural requirements of a manualized parent coaching program like Project ImPACT (i.e., providing constructive coaching comments), but the clinician's approach would be at low competency. Furthermore, it is possible that manual adherence may be related to a cognitive mechanism of consultation (e.g., proceduralization), while competency may be related to a skill mechanism (i.e., quality at which an intervention is delivered).

Manual adherence. Adherence to the Project ImPACT manual content was assessed by coders naïve to session number or timepoint via the Project ImPACT Coaching Fidelity Checklist (Ingersoll & Dvortcsak, 2019) to identify the extent to which clinicians adhered to lesson plans and content from the manual. Items include appropriate use of Project ImPACT program materials (e.g., use of handouts and the parent manual), appropriate description of Project ImPACT NDBI techniques to caregivers, sufficient time spent on critical Project ImPACT session components (e.g., at least 12 minutes of caregiver practice time), and presence of parent coaching clinical skills (e.g., provision of effective positive coaching feedback). Certain items that are considered more central to a parent coaching approach are weighted heavier (e.g., sufficient caregiver practice time, providing corrective feedback to caregivers), whereas other items are weighted less (e.g., assigning reading for the next session as homework). Clinically, a score of 80% or higher on this measure is considered “at fidelity.” All Project ImPACT sessions were randomly assigned to be scored by two independent coders naïve to timepoint and condition. Coders needed to meet training criteria of 3 videos in a row with 80% of items in exact agreement with the master coder prior to coding for the actual dataset. Monthly meetings with the coding team were held in which all members independently coded a video, and the group compared their scores to the master codes and discussed discrepancies. Once trained, coders met to discuss each videotaped session and resolved any disagreements by consensus. Consensus scores were used in the final analyses.

Parent coaching competency. Competency in parent coaching was assessed by coders naïve to session number or timepoint via the Parent Empowerment and Coaching in Early Intervention (PEACE) measure (Pellecchia et al., 2023) which utilizes a 5-point Likert scale to assess quality and competency in delivering collaborative coaching techniques used in parent-

mediated interventions (e.g., providing effective in-vivo corrective feedback to caregivers, problem-solving challenges with between-session practice). Subscales included: Collaboration, Demonstration, In-Vivo Feedback, and Reflection and Problem-Solving. Clinically, a score of 4 or 5 for each subscale would be considered “at fidelity.” As this measure is relatively new, no psychometric data has been published yet. All Project ImPACT sessions were randomly assigned to be scored for parent coaching competency by one primary coder naïve to timepoint and condition. Coders needed to meet training criteria of 3 videos in a row with 90% of items within one point of the master codes prior to coding for the actual dataset. Monthly meetings with the coding team were held in which all members independently coded a video, and the group compared their scores to the master codes and discussed discrepancies. Twenty-five percent of Project ImPACT sessions were scored by a second independent coder for reliability. Intra-class correlations ranged from 0.63 to 0.99, indicating moderate to excellent reliability; see Table 7 for all ICCs.

Table 7. Intra-class correlations for each PEACE subscale.

Subscale	Intra-class Correlation
Collaboration	0.69
Demonstration	0.79
In-Vivo Feedback	0.99
Reflection and Problem-Solving	0.63

Data Collection for Questionnaires

During the consultation period, clinicians completed weekly online questionnaires about the following: cancellations/rescheduling information, case penetration (Penetrability Formula), perceptions about Project ImPACT (Perceived Characteristics of Intervention Scale), and perceptions on the usability of that week's consultation training activity (Implementation Strategy Usability Scale). No clinician questionnaires were collected during the 8-week follow up period. After the follow up period, clinicians submitted a final recorded session and set of questionnaires.

Caregivers completed a caregiver-report questionnaire about autism-related challenges (Autism Impact Measure) via online questionnaire at 5 time points: at intake, at the end of each consultation condition (i.e., 4-, 8-, and 12-weeks into the consultation phase), and at the end of the 8-week follow up period.

Measures

Agency characteristics. Agency characteristics included the rurality of the agency's main office (as defined by the Rural-Urban commuting area codes maintained by the U.S. Department of Agriculture's Economic Research Service), implementation climate (Implementation Climate Scale), and attitudes towards evidence-based practices (Evidence Based Practice Attitudes Scale).

Implementation Climate Scale (ICS). Implementation climate refers to clinician perspectives on the extent to which their agency is supportive of evidence-based practice implementation. The Implementation Climate Scale (Ehrhart et al., 2014) is scored on a 1-5 Likert scale and examines the extent to which clinicians feel that their agency prioritizes evidence-based practices (Focus on EBP), provides training on evidence-based practices

(Educational Support for EBP), holds clinicians who are experts in evidence-based practices in high regard (Recognition for EBP), and provides financial incentives for clinicians who are implementing evidence-based practices (Rewards for EBP). Internal consistency on the measure is high, with Cronbach's alphas for all subscales ranging from 0.81-0.91 (Ehrhart et al., 2014).

Evidence-Based Practices Attitude Scale (EBPAS). The Evidence-Based Practices Attitude Scale uses a 1-5 Likert scale to measure clinician attitudes towards new evidence-based practices, including their openness to trying new interventions (Openness), the degree to which they believe that interventions developed in research settings are clinically relevant or useful (Divergence; higher numbers indicate higher levels of divergence or distrust), and their likelihood to implement an evidence-based practice if their agency required it (Requirements) or it appeared appealing to them (Appeal). Internal consistency on the scale is moderate to high, with Cronbach's alphas ranging from .59 to .90, with an overall scale alpha of .77 (Aarons, 2004).

Demographic information. Demographic information was collected at intake for clinicians and caregivers. Demographic information included race/ethnicity, age, gender, household income, highest level of education, and rurality of the home address (as defined by the Rural-Urban commuting area codes maintained by the U.S. Department of Agriculture's Economic Research Service). Clinicians also reported on their previous training experiences and certifications (e.g., Board Certified Behavior Analyst).

Case penetration. Case penetration was measured via weekly online questionnaires using the Penetrability Formula (Stiles et al., 2002), which uses clinician report of the total number of active Project ImPACT cases divided by the total number of eligible clients on their caseload (i.e., autistic youth under age 6) to yield a percentage. Scores ranged from 0-100.

Penetration was calculated separately for clients with Medicaid insurance and for clients with private insurance or self-pay.

Implementation Strategy Usability Scale. Clinicians reported on the usability of each consultation component weekly across the 4 weeks of each consultation condition using the Implementation Strategy Usability Scale (Lyon et al., 2021), which is an adapted version of the Systems Usability Scale (Brooke, 1996). This 10-item questionnaire uses a 5-point Likert scale to examine clinician perceptions about the complexity of using an implementation strategy (e.g., role plays). Likert scale ratings are used to compute a total usability score, which ranges from 0-100. Internal consistency on the Systems Usability Scale is high, with a mean Cronbach's alpha of 0.91 and a range of 0.83 to 0.97 (J. R. Lewis, 2018).

Perceived Characteristics of Intervention Scale. At baseline, across the consultation period, and at follow-up, clinicians completed the Perceived Characteristics of Intervention Scale (Cook et al., 2015) weekly about their perceptions of Project ImPACT. This 18-item questionnaire includes items related to various constructs from Rogers' Diffusions of Innovation theory (Rogers, 2010): Relative Advantage (i.e., the extent to which the intervention is more effective than other therapies the clinician has used), Compatibility (i.e., the extent to which the intervention fits well with the clinician's clinical judgment and preferences), Complexity (i.e., the extent to which the intervention is clear and easy to use), Trialability (i.e., the extent to which the intervention is easy to try out with clients), Observability (i.e., the extent to which the intervention produces improvements in clients that are easy to see), Potential for Reinvention (i.e., the extent to which the intervention can be adapted to fit the needs of clients and the treatment setting), Task Issues (i.e., the extent to which the intervention improves the quality of the clinician's work), Nature of Knowledge (i.e., the extent to which the knowledge and skills

needed to implement the intervention can be effectively taught), and Augmentation-Technical Support (i.e., the extent to which the intervention manual and supportive materials are helpful). All items are rated on a 7-point Likert scale and the overall PCIS score is established by computing the mean of all items. The PCIS is unidimensional in structure and has moderate to good reliability, with moderate to high internal consistency (Cronbach's alphas ranging from .67 to .95; Cook et al., 2015).

Autism Impact Measure. To assess for changes in child outcomes over time, caregivers completed the Autism Impact Measure (Kanne et al., 2014), which includes ratings of autism-related challenges in the domains of Communication, Social Reciprocity, Peer Interaction, Repetitive Behavior, and Adaptive Behavior. The Autism Impact Measure is sensitive to change in short-term autism interventions and has been used in trials of Project ImPACT (Mazurek et al., 2020). Internal consistency on the Autism Impact Measure is high, with the Cronbach's alpha for the total score being 0.96 and Cronbach's alphas for all subscales ranging from 0.79 to 0.91 (Houghton et al., 2019).

Data Analytic Plan

Visual Analysis. Visual analysis involves systematically evaluating data patterns to determine whether there is evidence of an observable, functional relation between an independent variable and a dependent variable. Advantages of visual analysis include the ability to evaluate data for individuals or small groups, utility in facilitating data-based research decisions (e.g., adapting the research design or intervention as data is collected), and the ability to evaluate treatment effects for individual cases or participants (using comparisons that account for an individual's baseline performance and subsequent changes in treatment phases; Barton et al., 2018).

Formative visual analysis is a systematic process to evaluate data patterns within and across conditions *during* the data collection process throughout the research project, allowing the researcher(s) to adjust the study design to ensure that participants benefit optimally from their involvement and/or to ensure optimal experimental control for data interpretation purposes (Barton et al., 2018, p. 180). This process was utilized throughout data collection, which helped the research team identify the methodological flaw with the artificially high baseline fidelity scores for sessions 1 and 2 that was described above.

Summative visual analysis is used to evaluate: a) whether there is a functional relation between the independent variable and dependent variable, and b) to assess the magnitude of the treatment effect (Barton et al., 2018, p. 181). Standards for single-case designs require a minimum of 3 demonstrations of “temporally-related and consistent behavior change” to qualify as evidence “with some reservations” for a functional relation (Barton et al., 2018, p. 181; Kratochwill et al., 2013). Summative visual analysis was used evaluate stability, level, and trend within phases. Stability refers to the presence of 3 to 5 baseline sessions with stable measurements of the dependent variable (i.e., observations do not vary widely across 3-5 consecutive sessions; Barton et al., 2018, p. 181). Level refers to the amount of behavior measured on the y axis and is typically described as low, moderate, or high (Barton et al., 2018, p. 181). Trend refers to the slope and direction for outcome measurements over time, including trend direction (i.e., accelerating, decelerating, or zero celerating/parallel to the x axis), magnitude (i.e., steep or gradual), and stability of the trend (Barton et al., 2018, p. 184). An overview the visual analysis approach for Aim 1 is presented below.

Visual Analysis for Aim 1: Effects of the group consultation model and each consultation component on clinician manual adherence and parent coaching competency. Only data from the

3 agencies with sufficient baseline data were used for Aim 1 analysis. A summative between conditions visual analysis was used to evaluate changes in data patterns (changes in level, trend, and variability from one condition to the adjacent condition; Barton et al., 2018, p. 190). We conducted a visual analysis for each agency (using weekly data averaged across all clinicians at the agency). Hypotheses about the presence of a functional relation were tentatively confirmed or rejected based on whether the consultation component was associated with a change in level from one condition to the subsequent condition. We hypothesized that consultation components would be associated with a change in level from condition to condition. Less emphasis was placed on trends and changes in trend direction because we did not have any hypotheses that scores would vary within a 4-week period due to the effect of time. Thus, change in level was the main indication of a significant finding; however, if there was a clear and meaningful shift in trend direction (e.g., steep decelerating trend to gradual accelerating trend but no change in mean level), this was also used when determining which phase changes were considered significant. This effect needed to be replicated across all 3 agencies (inter-case replication) for the functional relation to be tentatively confirmed; recommendations for interpreting single-case designs suggest that an effect should be replicated across at least 3 cases.

Ideally, a functional relation should be confirmed when there is evidence of both *intra*-case and *inter*-case replication (Barton et al., 2018). Resource limitations prevented our research design from demonstrating replicated effects within agencies (intra-case replication) – this would require multiple baseline phases and multiple conditions of each consultation component, resulting in a consultation period that would last well beyond the present 3-month duration and thereby make participant recruitment and retention unlikely. Therefore, any functional relations

that were identified in this study could only be labeled as “tentatively confirmed” if inter-case replication was found.

Statistical Analysis. An overview the statistical analysis used for each aim is presented in this section. Recent advances in statistical modeling have resulted in the specification of multilevel models that can provide effect sizes for each treatment element while also accounting for the specific features of single-case experimental designs (e.g., handling auto-correlation among consecutive datapoints, correcting for design-specific error structures). These advances have resulted in statistical packages in the R statistics platform (e.g., *sdhlm*). These tools were used to analyze results for Aims 1, 3, and 4. Specific model parameters are presented in the Results section.

Multilevel modeling for Aim 1: Effects of the group consultation model and each consultation component on clinician manual adherence and parent coaching competency. Only data from the 3 agencies with sufficient baseline data were used for Aim 1 analysis. We examined the effects of individual components of the group consultation model (i.e., consultation training activities) on clinician fidelity (i.e., manual adherence, parent coaching competency). To do so, we utilized 2-level multilevel models to estimate the change in level from the average baseline fidelity score to the average fidelity level for each consultation training activity condition (i.e., changes in level from baseline to each consultation condition). For example, the model estimated the magnitude of change from the Baseline level to the Feedback level, and whether the change was statistically significant. Specific model parameters for all models are presented in the Results section.

The models provide two sets of estimates: 1) the average effect of each consultation component on fidelity scores *averaged across clinicians* (i.e., results from the 2-level model),

and 2) *participant-specific* parameter estimates that demonstrate the effect of consultation components on an individual clinician's fidelity scores. First, estimates for the 2-level model (averaging across clinicians) include estimated fixed effects for each consultation component and random deviations for each clinician. These deviations are obtained from the random effect estimates, which included an intercept for each clinician and a random component for each of the 3 consultation phases. These random effects for the intercept allow the model to vary for each clinician in regard to their baseline fidelity level. The random effects for each consultation component allowed for the estimated effects of each consultation component to vary across clinicians (e.g., the effect of feedback could have a stronger magnitude for clinician 0201 than 0502). Next, the participant-specific parameters are estimated using the fixed and random effect estimates from the 2-level model. They are a combination of fixed effects and the between-participant variance. These participant-specific parameters quantify the effects of each consultation component on an individual clinician's fidelity levels. However, the model does not compute tests of statistical significance for effects of each consultation component on individual clinicians. In order to test whether the observed effects were statistically significant for a given clinician, one would need to fit separate multilevel models for each of the 5 outcome variables for each clinician (a total of 60 separate models), which was beyond the scope of this project. Also, those separate individual multilevel models would not account for between-participant variation, which is a major strength of the two-level multilevel models used in this analysis. Although we hoped to include a third level of nesting (i.e., agency-level effects such as how the rurality of agency #1 affects estimates), our sample size was too small to do so.

Aim 2: Relative usability of each consultation component. We examined statistical differences among mean ratings of each consultation component (i.e., training activity) on the

Implementation Strategy Usability Scale. Usability ratings were averaged for each consultation condition for each clinician and then examined using a repeated measures one-way ANOVA. Questionnaires from all 6 agencies were utilized for this analysis.

Aim 3: Effects of the group consultation model on case penetration and clinician perceptions of Project ImPACT over time. We used two-level multilevel models to examine changes in case penetration and clinician perceptions of Project ImPACT from the Baseline phase to the Consultation phase. Observations (i.e., weekly case penetration data, overall PCIS scores) were nested within clinicians. Questionnaires from all 6 agencies were utilized for this analysis. We ran separate case penetration models for clients using Medicaid insurance and clients using private insurance/self pay.

Due to the fairly homogenous small sample size and concerns about including too many predictors in the model, we did not include clinician demographics/previous training experiences in the multilevel model predicting perceptions of Project ImPACT on the PCIS. However, we were still interested in the relationship between these variables and PCIS scores. Thus, we examined the relationships between PCIS scores averaged for each clinician and a number of provider demographic, training, and caseload variables using Pearson's correlations for interval data and Spearman's rho for ordinal data.

Exploratory Aim 4: Effect of Project ImPACT on child social communication outcomes over time. We fit a series of two-level multilevel models to estimate the extent to which Project ImPACT was associated with changes in parent-reported child social communication outcomes over time. Observations (i.e., caregiver ratings on the Autism Impact Measure) were nested within each child. There were 3 multilevel models in total (Communication, Social Reciprocity,

and Peer Interaction). Questionnaires from families across 5 agencies with sufficient data were utilized for this analysis.