Using an Embedded Mixed Methods Design to Assess and Improve Intervention Acceptability of an Equity-Focused Intervention: A Methodological Demonstration

Eoin Bastable, PhD[1], Paul Meng, PhD[1] (iD), Sarah Fairbanks Falcon, PhD[1], and Kent McIntosh, PhD[1] (iD)

Abstract

Educators have shown reluctance to implement interventions aimed at improving racial equity in school disciplinary practice. Mixed methods were applied to assess and improve the acceptability of a new intervention designed to reduce racial disparities in school discipline. A descriptive concurrent parallel design was used to assess U.S. educators' perceptions of the acceptability of the intervention. Quantitative findings from professional development workshops introducing the intervention to 118 educators were corroborated with qualitative findings from a separate sample of 4 teachers who implemented it in their classrooms. Quantitative findings indicated that the intervention was acceptable to a broad range of potential implementers, and qualitative findings were used to modify the intervention to further improve its utility. The strengths, limitations, and implications of embedding mixed methods approaches to assess school-based interventions are also discussed.

A central challenge in education is not only developing potential solutions (e.g., interventions) to important problems but also assessing the extent to which potential solutions are acceptable to school personnel. Widespread evidence of racial discipline disparities in U.S. schools has underscored an urgent need for interventions that educators can implement to improve racial equity in student outcomes (Carter, Skiba, Arredondo, & Pollock, 2017; Skiba & Losen, 2016). However, individuals in U.S. society generally remain reluctant to address issues related to race or ethnicity (DiAngelo, 2011; Goff, Jackson, Nichols, & Di Leone, 2013). Educators, like the rest of U.S. society, are reported to be ambivalent or avoidant toward examining race and equity in schools (Bastable & McIntosh, 2019; Singleton, 2015; Tatum, 2017). Therefore, it is necessary to assess and improve how educators perceive the effectiveness and feasibility of equity-focused school interventions. We embedded mixed methods within the study design to assess and improve the acceptability of an intervention called ReACT, developed to reduce racial disproportionality in school discipline (McIntosh, Ellwood, McCall, & Girvan, 2018).

ReACT is a school-based professional development intervention designed to help educators improve racial

equity in school discipline practice. The ReACT intervention includes the following three elements: (a) training educators to assess discipline data to detect patterns of disproportionality (e.g., assigning more discipline referrals to Black students for defiance compared with other racial/ethnic groups), (b) assisting classroom teachers to adapt their school and classroom behavior support systems to be more culturally responsive (e.g., aligning school and home behavioral expectations), and (c) training educators on implicit bias and using strategies to neutralize implicit bias in school discipline decision-making. The intervention also includes ongoing coaching provided to school teams or individual educators.

Elements of ReACT were evaluated in two previous studies. A school case study documented reductions in discipline referrals assigned to Black students compared with White students in a K–Grade 8 setting (McIntosh et al.,

[1]University of Oregon, Portland, USA

Corresponding Author:
Eoin Bastable, 4520 NE 32nd Place, Portland, OR 97211, USA.
Email: ebastabl@uoregon.edu

2018). A single-case design study showed a functional relation between use of the intervention (implemented across four teachers) and increased equity in student–teacher interactions for Black students (Gion, McIntosh, & Falcon, 2019).

Methods of Assessing Intervention Acceptability

Instead of criticizing educators for implementing interventions with poor fidelity, it may be more helpful to examine an intervention's perceived acceptability by school personnel. Acceptability refers to whether potential implementers of an intervention, based on their knowledge or direct experience with the intervention, perceive it as agreeable or satisfactory (Proctor et al., 2011). In school settings, acceptability is often assessed after interventions are implemented with a rating scale, instead of beforehand with an eye to improving specific aspects of a new practice.

Acceptability is a multi-dimensional construct typically analyzed using data collected from different sources: surveys, key informant interviews, and focus groups. Mixed methods research offers a promising avenue for improving the understanding of intervention acceptability by assessing interventions from different vantage points (e.g., quantitative surveys of large groups and qualitative interviews with individual implementers). Increasingly, mixed methods are used to assess intervention acceptability by helping identify barriers and facilitators to implementation, or as a tool to refine interventions (Palinkas & Cooper, 2017). For example, Aarons, Ehrhart, Farahnak, and Hurlburt (2015) compared findings from analyses of qualitative and quantitative data to evaluate the acceptability of a leadership intervention to help staff implement evidence-based practices in mental health service agencies.

Due to the complex nature of improving racial equity in school discipline practice, scholars have called for more integrated methodological approaches to identify potential solutions (Carter et al., 2017; Klingner & Boardman, 2011). Specifically, analytic approaches that utilize both statistical analyses of quantitative data and in-depth qualitative interviews have been proposed (Skiba, Arredondo, & Rausch, 2014). Increasingly, methodological approaches from the field of Implementation Science (e.g., hybrid, step-wedge designs) have been considered to assess the effectiveness and utility of interventions developed for schools (Leeman et al., 2018; Lyon & Bruns, 2019). More flexible methodological approaches are needed to enhance the effectiveness and acceptability of school-based interventions aimed at reducing discipline disparities.

It is important to gain a more robust understanding of educators' willingness to adopt equity-focused approaches that may cause discomfort or raise defensiveness among school personnel (e.g., challenging educators' pre-conceived notions of fairness, equity, or neutrality). Understanding the perspectives of school personnel throughout the design, implementation, and evaluation stages of intervention development may contribute to the generalizability and usability equity-focused practices (Skiba et al., 2014). Given the benefits of using mixed methods to contribute to a broader line of research on discipline disproportionality, it is surprising that few studies have applied this integrative methodology (Fenning et al., 2011; Haight, Gibson, Kayama, Marshall, & Wilson, 2014).

Purpose

The purpose of this article is to demonstrate the use of mixed methods to assess and enhance the acceptability of a school-wide professional development intervention to reduce discipline disproportionality. Specifically, we integrated quantitative and qualitative data to evaluate an equity-focused intervention (i.e., ReACT). We delivered an overview of the intervention during a series of full-day workshops, then used a validated intervention-acceptability measure to identify overall acceptability among a sample of educators and tested for any differences in acceptability by race/ethnicity, gender, and U.S. geographic region of workshop attendees. In addition to the quantitative analyses, we used with a pragmatic interview approach to obtain rich information from teachers who actually implemented the intervention in their classrooms. Next, we used a mixed methods (i.e., concurrent parallel design) to analyze the extent to which results obtained regarding acceptability were consistent across workshop participants who were just learning about the intervention, which we corroborated with reports of teachers who had implemented ReACT in a school setting (i.e., classrooms). Specifically, we asked the following research questions:

RQ1. To what extent do educators rate the intervention as acceptable and feasible, and do ratings vary by (a) educator characteristics and (b) experience actually implementing it (quantitative data)?

RQ2. What variables do classroom teachers identify as enablers and barriers to implementation of the intervention (qualitative data)?

Method

Mixed Methods Approach

We used a descriptive concurrent parallel design (Creswell & Clark, 2011) to analyze the quantitative and qualitative data collected for this study. Utilizing this approach, we determined the study elements at the outset, then collected quantitative and qualitative data in a parallel manner, but

analyzed them independently. Next, quantitative and qualitative results were integrated to support an overall interpretation of how school personnel perceived the acceptability of the ReACT intervention.

There are compelling reasons for using mixed methods during intervention development. First, analyzing both quantitative and qualitative results independently may provide a more robust understanding of acceptability. Second, qualitative data were viewed as helping to corroborate quantitative results by providing additional information on the research topic based on participants' own words and experiences implementing the intervention. Third, mixing methods allowed for comparing quantitative and qualitative data to increase the legitimacy (i.e., validity) of the overall study's findings (Teddlie & Tashakkori, 2003).

Settings and Participants

Quantitative strand: professional development workshop attendees. Participants for the first research question were a convenience sample of educators and administrators from three U.S. states (one in the Midwest, one in the Northeast, and one in the South) who elected to participate in 1-day professional development workshops focusing on ReACT and its elements, delivered in 2017. Sites were selected as a deliberate sample to provide geographic diversity in attendees to assess group differences in acceptability by U.S. region. Of the 181 attendees across all three sites, 118 (65%) consented to participate in the study. According to self-report, participants were working in or supporting schools that were implementing school-wide positive behavioral interventions and supports (SWPBIS) with high fidelity (26%), some fidelity (42%), or not at all (10%). The remaining 21% either did not respond or indicated that this item did not apply to them. See Table 1 for descriptive statistics.

Qualitative strand: classroom teachers. We deliberately conducted qualitative interviews with classroom teachers who had first-hand experiences implementing the intervention. We used the data collected from the qualitative interviews as an independent credibility check for the quantitative survey data. The sample was obtained by asking administrators in two schools in a large urban school district to identify individual teachers who required additional support in equitable classroom behavior support. Although the teachers agreed to participate in the study, they did not seek out the training as the workshop sample did. In addition, they actually implemented the intervention (with adequate fidelity) before rating its acceptability, instead of simply learning about it. These participants were four general education teachers who were coached and implemented the ReACT intervention in their classrooms in an elementary and a K-8 school in the Pacific Northwest in 2018. The teachers were interviewed about their experiences after approximately 1

Table 1. Descriptive Statistics for Workshop Participants (*n* = 118).

Participant characteristics	*n* (%)	Acceptability *M* (*SD*)
Gender		
Male	24 (20)	5.36 (0.54)
Female	87 (74)	5.19 (0.47)
Race/ethnicity		
Black	26 (22)	5.29 (0.47)
White	79 (67)	5.42 (0.49)
Others	13 (11)	5.18 (0.40)
Region		
Midwest	30 (25)	5.15 (0.50)
South	56 (47)	5.22 (0.48)
Northeast	32 (27)	5.27 (0.48)
Role		
Administrator	23 (19)	5.20 (0.50)
Teacher	15 (13)	5.17 (0.56)
Student support	30 (25)	5.22 (0.42)
Other (e.g., coach)	25 (21)	5.33 (0.52)
No response	17 (14)	5.11 (0.46)
Years in education		
0–5	14 (12)	5.19 (0.37)
6–10	15 (13)	5.17 (0.53)
11–20	49 (42)	5.27 (0.46)
>20	33 (28)	5.20 (0.56)
Fidelity of SWPBIS (self-report)		
Fidelity	31 (26)	5.25 (0.53)
Partial	48 (41)	5.15 (0.47)
No SWPBIS	12 (10)	5.09 (0.45)

Note. Acceptability was measured using the Primary Intervention Rating Scale (Lane et al., 2009). SWPBIS = school-wide positive behavioral interventions and supports.

month of implementing the ReACT intervention. Teachers identified as White, Non-Hispanic (*n* = 2), White, Hispanic (*n* = 1), and Asian/Pacific Islander (*n* = 1); and 75% were female (*n* = 3). Teachers reported working in the field of education for an average of 6.5 years (from 1 to 17 years).

Measures

Acceptability. To assess social validity for the workshop attendees and four classroom teachers who implemented the intervention, we used the *Primary Intervention Rating Scale* (PIRS; Lane et al., 2009). The PIRS is a 17-item measure of overall intervention acceptability for school-wide behavior support interventions. The PIRS has been validated in the context of implementation research in school settings (α = .97 for elementary level). PIRS survey items are rated on a 6-point Likert-type scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). PIRS survey items assess different aspects of acceptability of school-based interventions (e.g., item 7: "*I would be willing to use this intervention in the school setting*").

Teacher interview protocol. To obtain rich descriptions of the four classroom teachers' implementation experiences, we used a semi-structured interview protocol and format (available from the first author) based on the Critical Incident Technique (CIT; Butterfield, Borgen, Maglio, & Amundson, 2009; Flanagan, 1954). CIT was used to identify categories that either enabled or hindered implementation of the ReACT intervention by teachers in their classrooms. The CIT method has proven to be especially useful for interpreting how incidents or experiences described by practitioners can inform how to improve current practices or policies (Butterfield, Borgen, Amundson, & Maglio, 2005; Flanagan, 1954). We used this protocol to elicit responses from each participant during one-on-one phone interviews to collect specific, observable, and replicable incidents (critical incidents [CIs]; Flanagan, 1954) to address the second research question. Flanagan (1954) stated that interviews should continue until exhaustiveness or redundancy in data occurs (i.e., the point at which participants mention no new CIs and no new categories are needed to describe incidents).

We asked participants to discuss what helped and hindered their implementation of the intervention in their classrooms adapting a common CIT interview format (Butterfield et al., 2009). Specifically, we analyzed responses to the following two questions: (a) *What were the important events (i.e., specific behaviors, examples, or observable happenings) that helped you to implement the ReACT intervention in your classroom or school?* (b) *What were the important events (i.e., specific behaviors, examples, or observable happenings) that hindered the use of the ReACT intervention in your classroom or school?*

Procedure

All study documents received approval from the University of Oregon Human Subjects Institutional Review Board. Recruitment for the workshop attendee participants took place at the start of each 1-day (i.e., 6 hr) professional development workshop provided by the fourth author. These workshops described ReACT and provided training and practice on the three intervention components (i.e., analyzing disaggregated discipline data, culturally adapting behavior support practices, and training on understanding and neutralizing the effects of implicit bias on disciplinary decisions). We offered attendees the opportunity to participate at the start of the workshop with provision of the surveys in a hard copy packet and a description of the study. Participants completed the acceptability survey at the end of the workshop. The classroom teacher interviewees were recruited as part of their participation in a small-scale trial of the intervention (Gion, McIntosh, & Falcon, 2019). Each

teacher implemented the intervention with fidelity, as measured through direct observation. All four teachers implementing the intervention participated in the interviews. PIRS administration and interviews began 1 week after the intervention trial concluded. We recorded and transcribed the four participants' responses to ensure that data were collected verbatim, as recommended from prior CIT studies (Andreou, McIntosh, Ross, & Kahn, 2015; Butterfield et al., 2009). Interviews lasted 45 to 75 minutes and were conducted over the phone with teachers after school hours.

Analytic Plan

Quantitative analyses. We first computed mean scores for PIRS rating scores to determine the overall ratings of acceptability and willingness across all workshop participants. We then analyzed survey responses by participant characteristics to determine the extent to which acceptability was consistent across demographic characteristics, including (a) gender, (b) race/ethnicity, (c) U.S. region, (d) educational role, (e) years in education (median split), and (f) perceived fidelity of implementation of SWPBIS (high, moderate, or none). For each characteristic, we conducted a separate one-way analysis of variance (ANOVA) to assess differences in acceptability. We used a Bonferroni-corrected α level for significance testing ($\alpha = .01$) to account for family-wise error (Huberty & Morris, 1989). We checked the data for the standard ANOVA assumptions and found no violations.

Qualitative analyses. We adhered to steps described by Butterfield et al. (2009) using CIT procedures to analyze participant interviews. First, we extracted CIs from four interview transcripts (conducted with four educators) that were associated with the "frame of reference" (i.e., what helped or hindered implementation of the ReACT intervention) for the study (Flanagan, 1954). Next, we identified patterns, themes, and differences among CIs to formulate categories with headings to summarize major themes. We then reviewed and coded the interview transcripts to determine the fit of the additional CIs into categories. Butterfield (Butterfield et al., 2009) recommended 25% as a minimum participation rate needed to form a viable category (i.e., a category should be noted by at least 25% of the participants). The threshold participant rate used for this study was 50% (i.e., two participants). If the threshold of 50% was not met for a proposed category, we considered combining smaller categories with those already formed (Butterfield et al., 2009).

Credibility checks. When all CIs were reviewed, coded, and placed in operationally defined categories, we initiated a series of credibility checks to determine trustworthiness, as

used in other CIT studies (Bastable & McIntosh, 2019; McIntosh, Kelm, & Canizal Delabra, 2016). Our credibility checks served as important quality indicators of this type of qualitative approach (Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005). Credibility checks included: (a) recording and transcribing all interviews for accuracy, (b) submitting one interview for independent review to ensure that the protocol was followed, (c) establishing intercoder reliability in the extraction of CIs and categories formed, (d) submitting categories to expert review (e.g., *did you find the categories to be useful?*), and (e) evaluating the categories for theoretical agreement.

Critical incidents extraction check. We recruited and trained an independent reviewer with a doctorate in Special Education to extract CIs from one randomly selected interview transcript. The independent CI extraction was compared with the extraction conducted by a member of our research team. Intercoder agreement (ICA) was calculated by dividing the total number of CIs extracted by the total number of unique CIs identified across both extractions. The percentage of agreement was 100% for CIs extracted.

Category coding check. We chose at random 25% of the CIs and asked a member of our research team (who did not conduct the interviews) to review category headings and operational definitions. For this credibility check, the reviewer was asked to match headings with operational definitions provided in an electronic PowerPoint file. We used Andersson and Nilsson's (1964) recommended criterion of 80% agreement or higher as benchmark for reliability. Initial ICA was 85%. With feedback from the reviewer, one category title was modified (Consistent Use of Praise was changed to Inconsistent Use of Praise) and one category definition augmented (we added coaching on alternative classroom behaviors to Coaching on Positive Behavioral Strategies) to improve the overall fit of category title and category descriptions. After these modifications, ICA was raised to 100% for the categories formed.

Expert check. We recruited two experts from the field of education, scholars versed in the study's topic area and aware of current practices used to address racial equity in school settings. We asked the experts to review the final category titles and definitions and respond to a set of questions about whether they found the categories appropriate, surprising, or useful (Flanagan, 1954). The experts were asked the following questions: (a) *Do you find the categories to be useful?* (b) *Are you surprised by any of the categories?* (c) *Do you think there is anything missing based on your experience?* The experts agreed that all the categories generated were useful and relevant. One expert suggested some slight wording changes for category definitions. Overall, the experts did not report that anything was missing or surprising from a review of the category headings and definitions.

Mixed methods data analysis. Following completion of the quantitative and qualitative analyses, a descriptive concurrent parallel design was used to assess the extent to which the quantitative survey results corroborated the findings of a qualitative study using structured interviews with classroom teachers who actually implemented the intervention. The analytic approach was concurrent because all measures and methods were determined before both survey and interview data collection took place. Furthermore, integration only occurred at the conclusion of the study (Teddlie & Tashakkori, 2003). Findings (i.e., categories) that emerged from the qualitative study were interpreted based on results from the quantitative study.

Findings

Quantitative Results

Of the 118 respondents, 105 (89%) completed some portion of the PIRS rating scale. Complete PIRS rating data were available for 95 (81%) of the workshop participants and the four implementing teachers. Cronbach's alpha for PIRS from the sample of workshop attendees ($\alpha = .92$) was excellent. Due to the constraints presented by the sample size, we used mean substitution to generate PIRS scores for analyses. Each participant's PIRS average was interpreted as an index of overall intervention acceptability.

Overall, workshop attendees provided consistently high ratings for intervention acceptability. The mean rating for the PIRS scale was 5.23 on a scale of 1 to 6, between "Agree" and "Strongly Agree." These results indicated that the intervention and its components were regarded as both acceptable and feasible to the workshop participants surveyed. As seen in the right columns of Table 1, mean values by participant characteristics for acceptability were similar across participant characteristics, with all subgroup mean values above 5.0, and there were no statistically significant differences by participant group (all *p*-values above the family-wise α value of .01). In other words, the intervention was rated as highly acceptable to workshop attendees, regardless of the individual characteristics we assessed.

To corroborate these findings and present a check against error introduced by ratings of acceptability without actually implementing the intervention, we compared these scores with the PIRS ratings from the four classroom teachers who implemented the intervention with fidelity. This group also provided consistently high ratings, with a mean PIRS score of 5.32 and no responses for any items below "slightly agree." These results were congruent across samples and support a finding of high ratings of intervention acceptability across individual characteristics and experience implementing the intervention.

Table 2. Categories Reported by Classroom Teachers ($n = 4$) Implementing the ReACT Intervention.

Type of critical incident	Categories	Number of CIs (% of teachers reporting)
Helping	1. Receiving Feedback on Use of Praise and Corrections	7 (75)
	2. Coaching on Positive Behavioral Strategies	5 (75)
	3. Defining and Offering Examples of Praise	3 (50)
	4. Conducting a Student Preference Assessment	2 (50)
Hindering	1. Inconsistent Use of Praise	8 (100)
	2. Lacking Personal Capacity to Implement	5 (75)
	3. Lacking Alignment with Existing Classroom Practices or Teaching Philosophy	3 (50)
	4. Competing School Priorities/Tasks	3 (50)

Note. A 50% participation rate ($>$ two participants) was the minimal level acceptable for category formation. CI = critical incident.

Qualitative Findings

The results of the CIT analysis identified four Helping and four Hindering categories. We used an iterative process that included multiple revisions to operationalize definitions and titles (i.e., adding, dropping, and modifying categories titles and definitions to fit CIs into categories). The final count included eight categories, encompassing 34 CIs. Table 2 displays the final categories sorted by Helping and Hindering CIs. The table includes the category title, total number of CIs, and the representation rate (percentage of participants that endorsed categories/total number of participants). The table is ordered hierarchically, largest to smallest, by participant representation.

Helping incidents. CIs that participants described as enabling implementation were coded into four helping categories: Receiving Feedback on Use of Praise and Corrections, Coaching on Positive Behavioral Strategies, Defining and Offering Examples of Praise, and Conducting a Student Preference Assessment.

Receiving feedback on use of praise and corrections. Participants reported receiving feedback helped to increase praise and decrease corrections delivered to students during classroom lessons. A coach provided verbal feedback and data reports (e.g., graphed data, counts, rates, and ratio) on the type and amount of praise or corrections observed by race of the students. Participants discussed the benefits of receiving immediate feedback on observed use of praise and corrections during classroom instruction. Participant 4 reported:

> It was good to have [the coach] observing me and then sending me, you know, my little feedback every night, to kind of read through and see . . . because you're teaching, and you don't always totally know what you're doing or saying because you're just doing it.

Activities in this category included monitoring rates of praise and corrections delivered, examining trends on use of praise/corrections by race of student, and educators adjusting use of praise based on the coach's feedback. Participants described valuing simple and interpretable data to track their classroom performance.

Coaching on positive behavioral strategies. This category refers to receiving coaching to support use of positive behavioral strategies in classrooms. Activities included meeting with a mentor outside of class to share ideas about teaching practices and receiving guidance to manage student classroom behaviors (e.g., reminders to restate school rules in positive, concise language). Participant 2 appreciated receiving coaching on implementing positive behavioral strategies: "I thought the check in after about a week was also beneficial to just kind of talk things through and how things were going with the [classroom] strategies. So that was just me and the coach."

Defining and offering examples of praise. This category refers to describing and providing examples of what use of praise can look and sound like in classroom settings. Activities in this category include providing specific examples of praise statements, operationalizing use of praise, and providing rationales for delivering praise under different classroom conditions (individual vs. whole-class instruction). For example, Participant 1 reported on the benefits of differentiating how she used praise with her students:

> Well, I think at the beginning I was kind of confused about like, what kind of praise I was supposed to be giving, so I asked [the coach] when we sat down together. . . he said, there are different kinds of praise, [one type] which is good for relational rapport type praise, but he was more focusing on, you know, the behavior and performance praise. So, I thought that was useful.

Conducting a student preference assessment. This category refers to a strategy used in ReACT to assess what types of acknowledgment (e.g., public verbal praise and acknowledgment ticket) students prefer or not prefer to receive for positive behaviors displayed in the classroom. Participants

reported on the benefits of understanding what acknowledgment their students desired:

> I thought they were all going say the goldfish or even the "wow" ticket. Many of them wanted just like, verbal praise and that's not something that I think that I would have guessed, you know, so that was super helpful. (Participant 3)

Activities in this category include using paper questionnaires, group discussions, and individual interviews to collect information on students' prior classroom experiences with praise and specific preferences for being recognized by educators for meeting classroom expectations.

Hindering incidents. Participants described four hindering categories that impeded implementation of the ReACT intervention to fidelity: Inconsistent Use of Praise, Lacking Personal Capacity to Implement, Lacking Alignment with Existing Classroom Practices or Teaching Philosophy, and Competing School Priorities or Tasks.

Inconsistent use of praise. This category refers to participants reporting their irregular use of praise as hindering implementation of the ReACT intervention to fidelity. Behaviors included using praise sparsely (even when coached to offer more praise), attending more to negative behaviors of students (instead of intentionally ignoring), experiencing pressure or fatigue when asked to increase rates of praise, and expressing doubt whether increasing rates of praise would positively influence student behaviors.

Participant 1 commented on the challenge of delivering more praise to students during classroom instruction: "I feel like I had to bounce around to get the numbers of praise in and to sit with some who struggle with, you know, some of the writing assignments to get them started." Participant 3 reported, "I've been struggling with correcting students versus like, focusing on the positive or like what [the coach] had said about restating the rule. And so that was really challenging."

Lacking personal capacity to implement to fidelity. This category refers to participants' lacking confidence or the ability to implement the intervention to fidelity. Within this category participants questioned whether they could accurately self-monitor their rates of praise and corrections or self-assess how equitably they were delivering classroom supports to students without external coaching (e.g., a colleague observing). Participant 1 related obstacles experienced while implementing the intervention representative of this category:

> I didn't have time to manage in my head how many times I've called on this White person once and this Black person, you know? So, I guess maybe the feedback was helpful, but it was hard like, I can't say I was managing it in my head and trying to make it all come out even.

Lack of alignment with existing classroom practices or teaching philosophy. This category refers to not implementing the intervention to fidelity due to perceived lack of fit with existing classroom activities or participants' classroom management systems. Activities in this category include discounting requests to increase praise based on personal reasons (e.g., "I'm a positive person"), raising concerns about how to taper/reduce high rates of praise to address unwanted behaviors, and difficulty providing praise due to pedagogical approaches. For example, Participant 1 remarked, "some of the lessons or activities that I had planned didn't allow for so much praise."

Participants also raised concerns that implementing the intervention did not always feel authentic or aligned with their personal teaching approach. Participant 2 described this hindrance, "I feel like I'm being, unreal with, you know, like it's too much positivity, and . . . sometimes I feel like it's almost too much and for me. . . I give my students praise [that] is very individualized." Participant 4 also shared concerns about adapting and sustaining use of the intervention in her classroom:

> I didn't want to get stuck on, you know, praising students for being on task for weeks and weeks and weeks. It's something that I wanted to move on from, um, but I don't know how that this strategy would have allowed me to do that.

Competing school priorities or tasks. This category refers to school events and classroom duties that interfered with implementing the intervention to fidelity (e.g., class activities, testing). Activities include meeting with the coach outside of class time and struggling to use the intervention within the class schedule. Participant 4 described how her classroom schedule was viewed as a barrier to implementing the intervention:

> It was like, during our math time and it was kindergarten . . . with any grade level there's so many things to do that sometimes when he [the coach] was coming [to observe the class] I'm like, I'm so sorry, but today we have this special lesson coming . . . we're not doing [the intervention] that today, you know?

Integration of Results

We next integrated the quantitative and qualitative stands to further assess intervention acceptability. Survey data (PIRS ratings) were corroborated with interview data collected from the four teachers who actually implemented the ReACT intervention with fidelity in their classrooms. Ratings of acceptability were uniformly high across all participant demographic groups. Those who actually implemented the intervention also rated it highly and

identified specific incidents that helped or hindered their implementation.

The qualitative findings provided an additional set of rich information indicating which components of the ReACT were perceived by classroom teachers as helping them to implement the intervention to fidelity (e.g., defining and teaching expectations, personalizing classroom acknowledgements, viewing disaggregated data, and classroom coaching). Qualitative findings also helped to identify barriers not described in the PIRS data, which showed uniformly high ratings across all items (including all four teachers strongly agreeing that they "would be willing to use this intervention in the school setting"). The four teachers were asked to describe barriers perceived as obstacles to implementing the intervention to fidelity. These barriers included having to provide high rates of praise to students, lacking personal capacity or confidence to implement the intervention to fidelity (without external support), and balancing competing school priorities (other duties assigned as teachers).

Discussion

Discipline disproportionality remains a vexing and costly issue affecting schools and students, without clear evidence-based practices to solve it. In addition, educators' reluctance to acknowledge and address racial school discipline disparities presents another obstacle to implementing viable solutions. Hence, it can be valuable to assess the acceptability and feasibility of equity interventions through a range of methodological approaches. We used mixed methods to obtain data from multiple participant groups and perspectives. Findings indicated high ratings of acceptability across all participant demographic groups, including a diverse sample of workshop participants and teachers who actually implemented the intervention. Moreover, implementing teachers identified specific incidents that helped or hindered their implementation, which we used in our efforts to improve the intervention.

Interpretation and Application of Primary Findings

Intervention acceptability. Embedding mixed methods into the study design allowed for a more robust analysis of how school personnel perceived the acceptability of the ReACT intervention shared in workshops or implemented in classrooms. It was encouraging to see that intervention acceptability, as measured by a validated quantitative measure, was strong (mean values above 5 on a scale of 1 to 6) for all workshop groups. There were no significant differences in ratings by gender, race, U.S. region, role, or years in education. Given the reluctance of some educators to implement equity interventions (Bastable & McIntosh, 2019; DiAngelo, 2011)

and regional variations in perspectives regarding racial disproportionality (Shaw & Braden, 1990), the strong acceptability indicates promise for ReACT. Moreover, the four teachers who implemented the intervention (and completed the PIRS after implementation) had mean scores slightly higher than those who only heard about it. This congruence across samples allows for stronger trustworthiness of the findings, albeit with the possibility of inflated scores in all samples due to social desirability bias.

Implementation enablers and barriers. The qualitative strand of the study yielded thick descriptions from the experiences from teachers implementing the intervention, beyond simple ratings of social validity. Qualitative data added a level of detail that provided useful information for improving components of the intervention. For example, individual teachers found the classroom coaching and feedback to be indispensable for supporting implementation of the intervention. This finding aligns with the existing research on the utility of individual coaching and performance feedback delivered to classroom teachers (Bradshaw et al., 2018; Gregory, Allen, Mikami, Hafen, & Pianta, 2015; Reinke, Lewis-Palmer, & Merrell, 2008). Such results are heartening because they point to individual coaching as a key enabler. However, the findings are also somewhat discouraging because individual classroom coaching is costly (in terms of resources required) and thus is rarely provided by coaches in practice (Bastable, Massar, & McIntosh, 2019).

Integration of qualitative data provided more detailed information on perceived enablers and barriers that could be used to enhance overall acceptability of the ReACT intervention. Enabling factors included implementation supports that are common to many school interventions (e.g., coaching, direct teaching with examples, performance feedback; Sanetti, Collier-Meek, Long, Byron, & Kratochwill, 2015). Likewise, barriers such as competing initiatives or lack of resources are common concerns for school personnel (McIntosh et al., 2014). Interestingly, participants identified consistent use of behavior-specific praise across the school day as a challenge to implement ReACT to fidelity. Participants described skepticism (e.g., more praise may not improve behaviors) and identified barriers (e.g., finding enough opportunities to deliver praise during lessons) that indicated additional strategies or coaching may be needed to ensure that praise is delivered equitably by educators implementing ReACT in classrooms.

Intervention improvement. Although the quantitative data indicated that the intervention was acceptable to a diverse group of implementers, we found the interview results to be valuable in refining the intervention in an effort to make it more likely to be implemented with fidelity. For example, relying only on changing teacher practices without attending

to the systems and contexts that encourage those practices places all of the responsibility for success on the individual teacher. Instead, given the participants' positive experience with coaching (e.g., Coaching on Positive Behavioral Strategies), we will emphasize coaching support (e.g., dedicating resources to classroom coaching) to capitalize on this helping variable (McIntosh et al., 2016). In addition, although increasing behavior-specific praise is an effective and acceptable approach (e.g., Defining and Offering Examples of Praise) for creating a positive classroom environment, we have emphasized additional, less-intensive strategies to build positive student–teacher relationships to complement the focus on increasing praise rates (e.g., greeting students at the door, micro-affirmations, student strengths, and interests surveys).

Limitations and Strengths

This study provided an opportunity to evaluate the benefits and limitations of embedding a mixed methods design within a larger intervention development project. Although there are benefits to applying mixed methods approaches to explore this topic area, there are also limitations. In fact, mixed methods projects are often subject to a larger set of limitations because they may be judged against quality standards of multiple research methodology traditions.

The high acceptability ratings in the survey component of the study may have been affected by social desirability bias, in which respondents felt compelled to provide higher ratings than they might otherwise, or they may provide high ratings because providing one in the context of a workshop is easier than actually implementing with fidelity. This limitation was mitigated to some extent in that the classroom teachers provided similar acceptability ratings and implemented the intervention with high fidelity.

The samples used for this study were non-random (i.e., convenience, purposeful) and therefore did not represent typical school personnel. The workshop participants were likely already supportive of equity-focused approaches based on their attendance at the workshops (and willingness to complete a survey). Consequently, inferences and generalizability of the results of this study are limited only to the educators sampled. However, because workshop participants and classroom teachers both viewed components of the ReACT intervention favorably, the meta-inference quality was likely higher than if the results has been contradictory (Onwuegbuzie & Johnson, 2006). There was also a wide discrepancy in the two sample sizes used in the quantitative and qualitative strands of the study (118 vs. 4). Onwuegbuzie and Johnson (2006) noted that sampling differences can make integration in mixed methods studies challenging and can threaten the validity or credibility of results generated. Sampling discrepancies can also affect the quality of meta-inferences drawn from the data gathered.

These limitations also provide an opportunity to reflect on elements of the study that we believe could be improved if we were to replicate it in another project. One key issue is that the sample, although racially and ethnically diverse, was limited to educators and administrators. We could widen the sample to include a broader range of stakeholders, including students and family members. Specifically, students could be recruited to describe specific intervention strategies that helped or hindered the development of positive student–teacher relationships. Likewise, family members could describe experiences related to family–school partnerships.

To address threats to internal validity in mixed methods designs, the use of CIT may be advantageous. A feature of CIT is reaching exhaustiveness when analyzing interview data gathered from study participants. Exhaustiveness is defined as the point at which participants mention no new incidents, or no new categories emerged or are needed to describe CIs (Butterfield et al., 2009, p. 270). Exhaustiveness has been described as a useful criterion to improve the quality of inferences and strength the validity of mixed methods studies in the absence of statistical sampling methods (Flanagan, 1954; Teddlie & Tashakkori, 2003). CIT may offer an approach to increase internal validity, even with discrepant sample sizes, by collecting data until exhaustiveness is achieved.

CIT may also strengthen inside–outside legitimacy as described by Currall and Towler (2003). Legitimation refers to ensuring findings or inferences based on results are credible, trustworthy, dependable, transferable, and confirmable. Onwuegbuzie and Johnson (2006) described inside–outside legitimacy as the degree to which a researcher accurately represents an insider's perspective (e.g., teachers directly implementing the intervention in their classrooms) and an outsider's perspective (e.g., research team). A distinctive feature of CIT is interviewing "insiders" (e.g., educators) to understand turning points or to gain insights for improving existing practices or policies (Flanagan, 1954). Using CIT as a methodological approach may be viewed as a tool for strengthening inside-outside legtimacy in mixed methods studies.

To assess credibility of findings, this study included five credibility checks, some conducted by trained reviewers (i.e., outsiders) to assess the content validity of data and categories formed during a study (Butterfield et al., 2009). The credibility checks built into CIT studies could help to address threats to validity (i.e., legitimacy) related to sampling or recruitment procedures. Furthermore, as a qualitative approach, CIT may improve the strength of meta-inferences from data generated using different methods.

Implications for Research

Use of mixed methods designs to improve school interventions shows promise as an approach to enhance our current understanding of discipline disproportionality and feasible

remedies to address this issue in schools. Mixed methods approaches are well suited for assessing intervention acceptability across a broad range of stakeholders in schools or other settings. Although mixed methods research requires substantial effort, such effort is warranted when considering the effort wasted in developing a potentially efficacious intervention that is not acceptable to school administrators or classroom teachers. The evidence demonstrating the effectiveness of the ReACT intervention has to date been limited to a few studies (Gion et al., 2019; McIntosh et al., 2018). It is currently too early to recommend widespread use of ReACT. However, there appear to be elements of this intervention that are acceptable to a diverse group of educators (e.g., across race, regions, and roles) that may make it appealing to other educators.

It is possible that the acceptability of the ReACT intervention could be improved by aligning the intervention to fit within existing school-wide frameworks rather than as a stand-alone intervention (Good, McIntosh, & Gietz, 2011). For example, classroom teachers described potential threats to implementing ReACT to fidelity that included a lack of alignment between the intervention and their teaching approach and implementing new strategies alongside competing school tasks or priorities. Although these types of hindrances may be common in school settings, such obstacles could be mitigated by helping teachers to adapt the intervention to fit their contexts and by ensuring school leaders prioritize disciplinary equity as school-wide goal. Overall, use of mixed methods not only advanced our current knowledge of this important topic area, but also helped us better understand what enabled or hindered key school stakeholders from implementing the ReACT intervention to fidelity.

Based on our experiences applying this intergrative approach, we plan to continue to embed mixed methods into our development project. For example, we will add a CIT interview study to complement our randomized controlled trial to assess implementation of the full, school-wide ReACT intervention. Use of a mixed methods design will allow us to assess modifications to the intervention and better understand how educators implement ReACT in a school-wide context. Such research could also reveal how to more effectively help educators to address school discipline disparities that are not captured by measures typically used in randomized controlled trials.

Despite the advantages of using mixed methods to study a topic like discipline disproportionality, there may be reasons why this approach is not used more frequently. Funding structures used to support education research (e.g., Institute of Education Sciences) generally prioritize quantitative approaches. Furthermore, researchers typically seek to publish separate studies (i.e., quantitative and qualitative) rather than combining methods in a single study. Although addressing such concerns is beyond the scope of this article, there is clearly a need to consider how to promote broader use of this methodological approach to advance educational research.

ORCID iDs

Paul Meng [iD] https://orcid.org/0000-0002-0278-2996
Kent McIntosh [iD] https://orcid.org/0000-0002-7765-2993

References

Aarons, G. A., Ehrhart, M. G., Farahnak, L. R., & Hurlburt, M. S. (2015). Leadership and organizational change for implementation (LOCI): A randomized mixed method pilot study of a leadership and organization development intervention for evidence-based practice implementation. *Implementation Science*, *10*, Article 11. doi:10.1186/s13012-014-0192-y

Andersson, B.-E., & Nilsson, S.-G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology*, *48*, 398–403.

Andreou, T. E., McIntosh, K., Ross, S. W., & Kahn, J. D. (2015). Critical incidents in sustaining school-wide positive behavioral interventions and supports. *The Journal of Special Education*, *49*, 157–167. doi:10.1177/0022466914554298

Bastable, E., & McIntosh, K. (2019). *Exploring educators' commitment to racial equity: A qualitative study of critical incidents*. Manuscript submitted for publication.

Bastable, E., Massar, M., & McIntosh, K. (2019). A survey of team members' perceptions of coaching activities related to tier 1 SWPBIS implementation. *Journal of Positive Behavior Supports*. Advance online publication. doi:10.1177/1098300719861566

Bradshaw, C. P., Pas, E. T., Bottiani, J. H., Debnam, K. J., Reinke, W. M., Herman, K. C., & Rosenberg, M. S. (2018). Promoting cultural responsivity and student engagement through double check coaching of classroom teachers: An efficacy study. *School Psychology Review*, *47*, 118–134.

Brantlinger, E., Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children*, *71*, 195–207. doi:10.1177/001440290507100205

Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A.-S. T. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, *5*, 475–497.

Butterfield, L. D., Borgen, W. A., Maglio, A.-S. T., & Amundson, N. E. (2009). Using the enhanced critical incident technique in counselling psychology research. *Canadian Journal of Counselling*, *43*, 265–282.

Carter, P. L., Skiba, R., Arredondo, M. I., & Pollock, M. (2017). You can't fix what you don't look at: Acknowledging race in addressing racial discipline disparities. *Urban Education*, *52*, 207–235.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. SAGE.

Currall, S. C., & Towler, A. J. (2003). Research methods in management and organizational research: Toward integration of qualitative and quantitative techniques. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral sciences* (pp. 512–526). SAGE.

DiAngelo, R. (2011). White fragility. *The International Journal of Critical Pedagogy*, *3*, 54–70.

Fenning, P., Pigott, T., Engler, E., Bradshaw, K., Gamboney, E., Grunewald, S., . . . Kato, M. M. (2011, August 4–7). *A mixed methods approach examining disproportionality in school discipline*. Paper presented at the Annual Convention of the American Psychological Association, Washington, DC. Retrieved from https://escholarship.org/uc/item/46c8z6z2

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358.

Gion, C., McIntosh, K., & Falcon, F. S. (2019). *Effects of a multifaceted classroom intervention on racial disproportionality*. Manuscript submitted for publication.

Goff, P. A., Jackson, M. C., Nichols, A. H., & Di Leone, B. A. L. (2013). Anything but race: Avoiding racial discourse to avoid hurting you or me. *Psychology*, *4*, 335–339. doi:10.4236/psych.2013.43A048

Good, C., McIntosh, K., & Gietz, C. (2011). Integrating bullying prevention into school-wide positive behavior support. *Teaching Exceptional Children*, *44*, 48–56. doi:10.1177/004005991104400106

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2015). The promise of a teacher professional development program in reducing racial disparity in classroom exclusionary discipline. In D. J. Losen (Ed.), *Closing the school discipline gap: Research for policymakers* (pp. 166–179). Teachers College Press.

Haight, W., Gibson, P. A., Kayama, M., Marshall, J. M., & Wilson, R. (2014). An ecological-systems inquiry into racial disproportionalities in out-of-school suspensions from youth, caregiver and educator perspectives. *Children Youth Services Review*, *46*, 128–138.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*, 302–308.

Klingner, J. K., & Boardman, A. G. (2011). Addressing the "research gap" in special education through mixed methods. *Learning Disability Quarterly*, *34*, 208–218.

Lane, K. L., Kalberg, J. R., Bruhn, A. L., Driscoll, S. A., Wehby, J. H., & Elliott, S. N. (2009). Assessing social validity of school-wide positive behavior support plans: Evidence for the reliability and structure of the Primary Intervention Rating Scale. *School Psychology Review*, *38*, 135–144.

Leeman, J., Wiecha, J. L., Vu, M., Blitstein, J. L., Allgood, S., Lee, S., & Merlo, C. (2018). School health implementation tools: A mixed methods evaluation of factors influencing their use. *Implementation Science*, *13*, Article 48. doi:10.1186/s13012-018-0738-5

Lyon, A. R., & Bruns, E. J. (2019). From evidence to impact: Joining our best school mental health practices with our best implementation strategies. *School Mental Health*, *11*, 106–114.

McIntosh, K., Ellwood, K., McCall, L., & Girvan, E. J. (2018). Using discipline data within a PBIS framework to enhance equity in school discipline. *Intervention in School and Clinic*, *53*, 146–152.

McIntosh, K., Kelm, J. L., & Canizal Delabra, A. (2016). In search of how principals change: A qualitative study of events that help and hinder administrator support for school-wide PBIS. *Journal of Positive Behavior Interventions*, *18*, 100–110. doi:10.1177/1098300715599960

McIntosh, K., Predy, L. K., Upreti, G., Hume, A. E., Turri, M. G., & Mathews, S. (2014). Perceptions of contextual features related to implementation and sustainability of school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *16*, 29–41. doi:10.1177/1098300712470723

Onwuegbuzie, A. J., & Johnson, R. B. (2006). Special issue on "new directions in mixed methods research." *Research in the Schools*, *13*, 48–63.

Palinkas, L. A., & Cooper, B. R. (2017). Mixed methods evaluation in dissemination and implementation science. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (2nd ed., pp. 335–353). Oxford University Press.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 65–76.

Reinke, W. M., Lewis-Palmer, T. L., & Merrell, K. W. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*, 315–332.

Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C., Byron, J., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of School Psychology*, *53*, 209–229. doi:10.1016/j.jsp.2015.03.002

Shaw, S. R., & Braden, J. P. (1990). Race and gender bias in the administration of corporal punishment. *School Psychology Review*, *19*, 378–383.

Singleton, G. E. (2015). *Courageous conversations about race: A field guide for achieving equity in schools*. Corwin Press.

Skiba, R. J., Arredondo, M. I., & Rausch, M. K. (2014). *New and developing research on disparities in discipline* (The Equity Project). Bloomington: Indiana University. Retrieved from http://www.njjn.org/uploads/digital-library/OSF_Discipline-Disparities_Disparity_NewResearch_3.18.14.pdf

Skiba, R. J., & Losen, D. J. (2016). From reaction to prevention: Turning the page on school discipline. *American Educator*, *39*, 4–44.

Tatum, B. D. (2017). *Why are all the Black kids sitting together in the cafeteria? And other conversations about race*. Basic Books.

Teddlie, C., & Tashakkori, A. (2003). *Handbook of mixed methods in social behavioral research*. SAGE.