8

Commentary: Contextualizing Alternatives to RCTs: Measuring the Impact of a Non-Governmental Sponsorship Organization's Projects to Strengthen Children's Supportive Environments

Darcy L. Strouse (iD)*, Kathryn Moore*

Abstract

ChildFund International (ChildFund) is a child-focused International Non-Governmental Organization (INGO) which, since 1938, has worked with local implementing partners (LIPs), government, and other partner organizations to help create the safe environments children need to thrive. The purpose of this commentary is to reflect on the utility and possible application of the suggestions and study designs in this special issue to real-life intervention studies in dynamic context settings. The commentary provides three regional case examples with evaluation study lessons learned from ChildFund's global evaluation work. Specific insights from this issue are discussed regarding their application to effectiveness studies for projects delivered by a child-focused INGO, like ChildFund, in resource-limited settings. © 2019 The Authors. New Directions for Child and Adolescent Development published by Wiley Periodicals, Inc.

Child-focused development organizations dedicated to empowering communities to address and resolve complex social challenges, such as violence against children, face a dilemma with regard to

conducting rigorous impact evaluations of their programs. Particularly, the use of randomized controlled trials (RCTs) comes with many challenges, including: (a) ethical issues, such as the need to serve all population members at risk which prevents the use of a counterfactual in a program's evaluation; (b) timing issues, such as the urgency to act quickly and flexibly to deliver—and measure the impact of—an intervention response to an immediate risk or identified time-sensitive programmatic need, thus, there is often not sufficient time to plan and execute a RCT or other rigorous design; (c) financial and/or technical resources, to implement the often resource-intensive RCT; and (d) generalizability issues, such as implementing a rigorous impact evaluation that will yield results that cannot be extended beyond the experimental study context. Issues such as these have led to a call across the education intervention, human services programming, and international development literature to continue examining the suitability of alternative research designs for investigating the impact of interventions (e.g., Deterding & Solmeyer, 2018; Peck & Goldstein, 2016; Wynn, Dutta, & Nelson, 2005).

Conducting research to determine program impact in vulnerable settings further compounds the dilemma of conducting rigorous research studies. Such settings include populations dealing with extreme poverty; disease such as HIV and AIDS; geographic remoteness with limited or no access to government-mandated social services; situations of crisis and conflict; violence; and perpetuation of harmful traditional, cultural, and religious or gender-based norms, which further compounds the dilemma of conducting rigorous research studies. These challenging contexts, often referred to as *dynamic contexts*, are complex and diverse community settings in low- and middle-income countries comprised of vulnerable populations. Factors that make these contexts "dynamic" (and difficult for the implementation of resource intensive RCTs) include environments where social services and programs may be actively "intertwined and overlapping, provided by multiple advocates," and that such interventions exist in "complex community contexts" (Goodman, Epstein, & Sullivan, 2018, p. 60). These contexts may include populations that are on the move, not economically stable, have experienced trauma, and are influenced by local politics, culture, and service and programming access (Goodman et al., 2018).

The collection of papers in this special issue share experiences with the conduct of rigorous study of intervention impact and include both alternatives and enhancements to the use of "gold standard" RCTs. The purpose of this commentary is to reflect on the utility and possible application of the suggestions in this special issue to real-life intervention studies in dynamic settings. Specifically, we will reflect on how the insights in this issue can be implemented in studies on the effectiveness of programs provided by a child-focused international non-governmental organization (INGO) (i.e., ChildFund International) in resource-limited settings in Africa, the Americas, and Asia.

ChildFund International's Evaluation Practices and Organizational Considerations Informing Evaluation Study Design and Conduct

ChildFund International (ChildFund) [www.childfund.org] is a child-focused INGO which, since 1938, has worked with local implementing partners (LIPs), governments, and other partner organizations to help create the safe environments children need to thrive. The organization's twofold purpose is to (a) help deprived, excluded, and vulnerable children around the world. This is done by improving their lives and helping them become adults who bring positive change to their communities; and (b) promote societies that value, protect, and advance the worth and rights of children. ChildFund has direct programming in communities in twenty countries in Asia, Africa, and the Americas, including the United States. In 2018, our social service programs reached over 10.5 million children and family members. At ChildFund, we work across three key life stages to tailor our interventions to meet children's evolving needs as they develop from infancy through young adulthood.

Our Life Stage Approach (see ChildFund, 2018a) is comprised of a theory of change per "stage," outlining conditions that need to be met to support children reach a set of core developmental outcomes. These core outcomes are: (a) "Healthy and Secure Infants" for Life Stage 1 (ages 0 to 5); (b) "Educated and Confident Children" for Life Stage 2 (ages 6 to 14); and (c) "Skilled and Involved Youth" for Life Stage 3 (ages 15 to 24). This Life Stage Approach guides our evaluation efforts and "provides a framework for addressing the changing problems children face as they mature—whether at home, school or in the community—including the different forms of abuse, neglect, exploitation and violence" they may experience growing up in dynamic contexts (ChildFund, 2018a, p. 15).

In addition to the use of a global Monitoring and Evaluation (M&E) system that captures routine, standardized child protection and outcome data from all the countries and communities in which we work, we develop and implement a range of different methods for evaluating our programs in their specific contexts. For example, during the past decade, these evaluation studies have included rapid ethnographic studies (intensive 1 week to 10-day community case studies that involve qualitative data collection with community members, including stakeholders, caregivers, and children) to explore how violence and child protection systems impact children in specific communities (ChildFund, 2018a); quasi-experimental studies of a school-based violence prevention program in Central America (ChildFund, 2017); and a preschool-based disaster risk reduction (DRR) program in Indonesia (Proulx & Aboud, 2019), as well as numerous evaluation studies without counterfactuals (e.g., one-group pretest posttest study designs).

To date, organizational considerations that have informed the types of impact evaluation studies we conduct reflect on many of those

considerations identified in the extant literature, specifically: (a) ethical issues related to working with children and their families, (b) organizational technical resource and evaluation capacity considerations for implementing rigorous study designs, (c) the methodological requirements of RCTs and other experimental designs, (d) potential safety risk for evaluation support staff and study participants, (e) funding needs for RCTs and other rigorous study designs. An additional consideration is the need to maintain positive relationships with the communities in which we work due to the traditional sponsorship model of the organization. These are the communities in which our affiliated LIPs are entrenched and have a long-term commitment to communities—and the strength and success of programming is based on maintaining these relationships.

In addition, ChildFund has a "decision tree" process for determining the research cases in which rigorous impact studies, such as RCTs, should be considered and the extent to which organizational resources may be deployed to support the conduct of these studies. This approach includes the use of decision trees for both strategic or innovative programming projects (financed with traditional sponsorship funds) and grant-funded projects. For both decision tree processes, critical decision-point questions center on whether the guiding evaluation question (or proposed idea): (a) is related to the organization's core programming or another current strategic effort (strategic fit); (b) has been tested previously by ChildFund or others (contribution); and (c) reflects the study of service programming that is well-designed, well-established, and, thereby, ready to be evaluated (credible program design and readiness). Considerations for grant-funded projects also include whether the funding is from a donor with whom ChildFund is cultivating a long-term relationship based on evaluation results and if there will be sufficient funding available to use (and/or match) for rigorous evaluation.

Case Examples From Africa, the Americas, and Asia

To illustrate how we address and balance the pursuit of evaluation rigor with community service and relationship needs to generate community-based programmatic evidence in dynamic contexts, we share a set of case examples for evaluation studies we conducted in each of the regions in which we implement human service programming—Africa, the Americas, and Asia. Each of these case examples focuses on one of our Life Stage projects aimed at creating a safer world and improving developmental outcomes for children. For each of the three case examples we share (a) a brief overview of the intervention project, its evaluation study, and contributions and (b) highlights of our key experiences and lessons learned in designing and executing the evaluation study in a dynamic context setting.

The "Assuring the Essentials of Optimal Development for Infants and Young Children Affected by HIV and AIDS" Project in Kenya and Zambia

As part of our Life Stage 1 (ages 0 to 5) programming in Eastern and Southern Africa, ChildFund, supported by the Conrad N. Hilton Foundation, implemented a project focused on responsive caregiving and early stimulation known as "Assuring the Essentials of Optimal Development for Infants and Young Children Affected by HIV and AIDS" from January 2016 to July 2018. There is a high prevalence of HIV and AIDS in both Kenya and Zambia, and a lack of access for families to mandated services (e.g., health, education, nutrition, and social welfare). The aim of the project was for children aged 0–5 years in communities affected by HIV and AIDS to meet their developmental milestones by increasing the capacity of their caregivers to be responsive to children's physical, social, and emotional needs. ChildFund Kenya and ChildFund Zambia and their LIPs provided technical training in various aspects of Early Childhood Development (ECD), now known as nurturing care, through group parenting sessions and/or home visits. Training content included key messages and interactive activities on responsive caregiving, health, nutrition, child protection, stimulation/early learning, and safety/positive discipline knowledge and practices. The project also aimed to strengthen local government stakeholders' capacity to integrate parenting content on these issues into their existing or forthcoming policies, plans, and activities/services across sectors to support sustainability after the project ended. Project reach included thirty-six mentors (twelve in Kenya, twenty-four in Zambia), 347 facilitators (203 in Kenya, 144 in Zambia), and 5,601 caregivers (2,701 in Kenya, 2,900 in Zambia).

A one-group pretest-posttest study design was used to evaluate the community-based project across the two country contexts. The selection of a study design for this project was primarily based on intervention design, development, and delivery for this dynamic context. Mixed-method data collection at baseline and endline (23 months post-project start-up) included quantitative data on caregivers' socioeconomic and demographic characteristics; knowledge and practices related to responsive parenting and early stimulation; and barriers to and enablers of access to early childhood-related services. These data were collected through household surveys with caregivers and included measurement of caregiver: self-care; access to early childhood-related services (e.g., early learning and nutritional support); and level of play and communication with their children. Qualitative data were collected through focus groups with caregivers and key informant interviews with caregivers, facilitators and mentors, county and district government officials across health, education, social services, and ChildFund and LIP staff. Qualitative data were collected on the following topics: community's understanding of responsive caregiving and ECD; influence of the home and/or group parenting sessions on caregiver knowledge, skills,

attitudes, and practices; identification of barriers and enablers for accessing early childhood-related services; and the extent to which responsive care-giving, early stimulation, and supervision-based activities were integrated in the formal government system. A process evaluation study component assessed barriers and facilitators to project implementation.

Due to the nature of the nonexperimental, one-group pretest posttest design, changes in attitudes and practices among the program participants could not be attributed to the project with any certainty; however, at posttest, positive changes in outcomes for caregivers and their children were observed in the caregiver survey results. Improvements in caregivers' prac-tices were seen in the areas of play and communication and strengthened positive and protective relationships with their children. For example, there was a significant decrease both in physical violence (from 55% at baseline to 28% at endline) and verbal discipline (from 45% at baseline to 12% at endline). In addition, a health-linked outcome of the study was that care-givers also reported (via focus groups) that engaging with their children in communication and play served the important role of helping them observe their children's health, allowing the caregiver to know whether the child was sick or not. The project also contributed to positive policy outcomes. LIPs, subnational government stakeholders, and community facilitators reported improvements in their understanding of, and how, to effectively disseminate information and/or advocate for ECD. In Zambia, for example, as the coun-try prepared to have the first policy debate on a draft ECD policy, most par-ticipating government departments across sectors at the subnational level asserted that project participation influenced their interest to contribute to the policy based on their new understanding of the importance of nurturing care.

Given the dynamic nature of the community contexts for this project in both Kenya and Zambia, a context-related issue for the study that enhanced threats to internal validity (e.g., history, maturation, testing) was that during the project duration, other development and/or government organizations were implementing, competing, and/or supplemental interventions in the same communities. Although a process evaluation component was included in this study, it did not capture if and how study participants (caregivers) participated in other co-existing interventions during the study time period (e.g., the frequency of their child's participation in monthly government-provided growth monitoring and health promotion visits). Thus, the use of a one-group pretest posttest design, without these process data points made it difficult to determine if children's developmental status was linked to ChildFund's project, other outside services, or a combination of both.

In future studies of this intervention approach, addressing the above evaluation challenges will enable ChildFund to better understand which intervention strategies and/or combination(s) of strategies are most or equally effective in contributing to desired aims for enhancing caregiver capacity and supporting infants and young children in reaching

developmental milestones in safe, protective and playful environments. Our overall lessons from this evaluation study include the need to: (a) explore and apply the use of more rigorous study designs (at a minimum quasi-experimental) to these dynamic context settings to permit the analysis of project attribution; (b) explore research partnerships between in-country university-based researchers and international research advisors to support the conduct of more resource intensive study designs—and for jointly conducting all phases of the evaluation to ensure quality of study design and methods, adequate contextualization of tools, fidelity of methods employed, and quality of analysis and reporting; and (c) place an increased emphasis on implementation research in the evaluation design which includes adding a more rigorous process evaluation component to the overall study design to more comprehensively assess linkages between key outputs and outcomes. Future evidence from more rigorous studies, such as interrupted time series and regression discontinuity (RDD) designs, may be used to both showcase value added of a multi-sectoral coordination approach for nurturing care at subnational and community levels as well as support government, development partners, and funders to make crucial decisions on which interventions are best suited to different operating contexts within and across countries.

The "PUENTES Project" in Honduras

As a second case example, we share our experience designing and implementing a pilot study of school-based violence prevention programming in high-risk settings in the Americas. Specifically, to address significant issues of violence affecting our Life Stage 2 children, ages 9 to 13, in Honduras, ChildFund Honduras received support from the United States Agency for International Development (USAID) from September 2015 to February 2017 to implement the "Parents and Teachers Joining Forces for Children through Social Spaces (PUENTES)" project (Puente = Bridge). The main aim of the project was to pilot a comprehensive, curricular school-based violence prevention approach targeting violence in the home and at schools called Miles de Manos (MdM) or "Thousands of Hands." MdM, which consists of three components—family, school, and integration or "bridge"— that had previously been tested by ChildFund Honduras through its LIP in rural areas. To examine the impact of the approach in urban settings, PUENTES was delivered in three urban areas with some of the Honduras' highest crime and violence rates: Tegucigalpa, San Pedro Sula, and La Ceiba. Based on the heightened need to reduce violence against children in these violent urban communities, our pilot project sought to: (a) reduce the level of victimization reported by students, teachers, and parents; (b) increase the perceptions of security reported by students, teachers, and caregivers; (c) reduce the levels of aggression reported by students, teacher, and parents; and (d) increase the year-long attendance rate of 4th, 5th, and 6th graders.

During its 18-month implementation period, the project benefitted thirty-six schools, 265 teachers, and 2,269 families in these high-risk areas.

A quasi-experimental study design was used to evaluate the intervention in an urban, high-risk setting. This study design was selected for the project based on funding agency preferences, ChildFund's LIPs, and local context priorities and considerations. The baseline and final evaluation were conducted by consulting firms and groups of researchers outside of the ChildFund implementing organization (ChildFund Honduras). Seventy-two schools were included in the study (thirty-six intervention and thirty-six control schools). The intervention and control schools were required to have similar characteristics based on the following variables: socio-economic status, school size, geographical area, household composition, level of violence, and presence of (or lack of) other potentially confounding projects with similar aims.

Data collection using both quantitative (surveys with students, caregivers, and teachers) and qualitative (focus groups with participating teachers and caregivers and nonparticipating caregivers) methods was conducted at baseline and endline (18 months post-project start-up). Final evaluation results were promising as students in intervention schools reported feeling more safe and secure both in school and at home and caregivers reporting increased positive communication in the home and increased positive discipline in the home. However, on specific indicators, there were significant improvements for control schools in comparison to intervention schools, which may have been due to several challenges, both in respect to the design and execution of the study design.

One of the major design issues for this project centered on the project's school selection criteria—and the necessity to balance programing needs and ethical considerations (e.g., serving the schools where the MdM approach was most needed and identifying the participants most likely to benefit from the project) with study design rigor (e.g., high-need schools being assigned as control schools). An important consideration for the selection of the thirty-six intervention schools was to select schools where the MdM approach was most needed and the participants most likely to benefit from the project. Pre-baseline, when determining the school selection for comparable schools for each study arm, ChildFund Honduras, USAID, and the Honduras Ministry of Education agreed on selection criteria for intervention and control schools based on school size, enrollment rates, levels of violence, safety, logistical feasibility, presence of complementary interventions of USAID, and willingness to work within the project. After the application of this criteria, several schools with high need (high levels of violence) were found to have low enrollment rates and could have been screened out of participating. Instead, these schools were included in the study. Their contextual issues paired with the extant evidence suggesting that violence has an impact on enrollment and attendance rates indicated that the MdM approach might be most needed in these schools.

A second design issue was the need for more process evaluation of the study itself. A strength was that the project's monitoring and evaluation plan had built in monthly process monitoring of program fidelity. Observation tools were used by facilitators and the study's Monitoring and Evaluation Specialist to ensure frequent monitoring of the intervention and ensure quality and fidelity to the intervention. In addition, ChildFund technicians monitored every other program session facilitated by parents and teachers for at least 1 hour at each school. What was not included in the monitoring and evaluation framework was a process assessment for the implementation of the evaluation (e.g., school selection, data collection, and data analysis), which was further compounded using different evaluators for baseline and endline. Schools with other violence prevention, mass media, or other related interventions were not well-documented nor monitored over time. This had implications for study findings, particularly in limiting the ability for results to be attributed to the project intervention approach and producing confounding results.

A primary context-related challenge that influenced the conduct of the evaluation study was the influence of safety and security issues in the high-violence community settings for the project. Safety and security training was conducted at the outset of the PUENTES project with related safety protocols and informed consent procedures established. Risks were not systematically reassessed over time leading to later security threats to project and evaluation stakeholders, as well as threats to data quality. For example, ChildFund Honduras staff asserted that school directors had to "negotiate" with influential gang leaders to implement project and evaluation activities. Also, in one community a gang member stood in the door of the classroom when students were completing individualized surveys at follow-up. The latter could be considered as a form of intimidation, a significant safety and security risk to the students and entire school community and threat to the validity of the data.

In considering these challenges, ChildFund had some distinct lessons learned, particularly regarding the conduct of pilot studies with more rigorous designs, such as the PUENTES project's quasi-experimental design. First, as with our first case example project evaluation, there is a need for increased emphasis on understanding the intervention context such that contextual questions are added to the evaluation and project monitoring frameworks to better understand if, how, and for whom external (possibly competing) concurrent interventions are being conducted to identify confounding variables as they relate with study results. Second, during the life of a study, it is important to increase capacity for ongoing risk assessment and prioritizing a preventive and response approach to addressing safety and security risks. This includes systematic, external risk assessments conducted during the study design phase and at other points throughout implementation to better understand how external risks to the participants affect the school community (directors, teachers, students and their

families), project, and evaluation activities and the ultimate achievement of quality data and results. Third, an increased emphasis on process evaluation should be folded into the overall study design to better understand linkages between outputs and outcomes and to inform necessary changes in project activities and scope during the implementation period itself.

As ChildFund moves forward in scaling up the MdM approach for delivery as a part of a more comprehensive, global program model addressing school-based violence for Life Stage 2 children, we are using these lessons in designing pilot evaluation studies to examine the program model's contributions and attributions across a range of dynamic contexts including school communities in Sri Lanka and the Philippines in Asia, and Sierra Leone and Kenya in Africa.

The "Combatting Child Trafficking Through Sustainable Livelihood Development Project" in India

Our third case example highlights one of our community-based studies conducted in Asia of a project focused on creating opportunities for children as an alternative to exploitative labor. In India, due to high poverty levels, annually, thousands of children from Rajasthan's Udaipur, Dungarpur, Banswara Districts are trafficked across the border into districts of North Gujarat to work in cotton fields in extreme and adverse conditions. To counter this, working across our Life Stage 2 and Life Stage 3 with children aged 7 to 18 years, ChildFund India, with support from the Government of Germany, implemented the "Combatting Child Trafficking through Sustainable Livelihood Development" project in fifty villages across Udaipur District, covering a population of about 20,000, including approximately 8,000 children.

The aim of the three-year project was to utilize a community-based approach to raise awareness of child trafficking and associated protection risks, as well as to work with local government officials, community leaders, children and their caregivers to establish an enabling environment to prevent trafficking and increase the quality of life of children and adolescents vulnerable to trafficking. The project also worked with LIPs to empower local community and government actors to take coordinated action on child trafficking. The main interventions were awareness raising activities targeting all stakeholders at the community level (government and development partners to children and their caregivers); facilitating capacity building workshops on identifying and addressing child protection and trafficking risks specifically targeting local police, lawyers, development partners at the local level, government, and other local administrative staff; and conducting vocational technical training to support alternative livelihoods for vulnerable adolescents and their families. During its three-year duration, the project was able to provide technical training to 374 adolescents to pursue alternate livelihoods.

A one-group pretest-posttest study design was conducted to examine the contribution of the initiative within the Southern Rajasthan community context. Focus groups and key informant interviews were conducted at baseline and endline to collect qualitative data from key stakeholders such as government officials, community leaders, caregivers, and children (aged 10–16) participating in clubs. Quantitative data were collected at baseline and endline (36 months follow-up) through household surveys with caregivers and children and adolescents aged 10 to 16 years. Outcomes measured included: the percentage of children out of school; percentage of and reasons for children engaging in child labor; percentage of children and reasons for children engaging in cotton work; degree of parental consent for child's engagement in work; and parental awareness on child labor, trafficking, and child protection. Convenience and purposive sampling methods were used for the qualitative data collection. Stratified random sampling was utilized for collecting the quantitative data.

Although, due to the nonexperimental nature of the study design, project impact (attribution of outcomes to the intervention) could not be determined, results from the initiative were encouraging, reflecting an overall decrease in child trafficking trends, as well as increased awareness of child protection among children and their caregivers. Findings also revealed increased ownership of government officials and local development partners for preventing and responding to instances of child trafficking at the local level and children successfully engaged in alternative livelihoods.

Evaluation issues included resource constraints and community participation challenges. Specifically, resource issues such as the time needed for the study design and planning stage (including the time needed to sample and survey a control group), contributed to the decision to conduct a pretest-posttest study for assessment of the project over more rigorous, quasi-experimental, or experimental study design options. The evaluation team attempted to add rigor to the study by utilizing random sampling for the quantitative data collection at posttest. However, the more rigorous approach to sampling took careful thought and more time planning. Thus, random sampling of villages was ruled out in discussion between the evaluator and project staff because it was observed that villages close to the "headquarters" or "block" location typically have greater access to government officials and information, thus results would likely be skewed positively if including respondents from these areas. As an alternative, stratified sampling was employed (strata included population of the village, presence of various tribal sub groups, distance to headquarters, distance to destination site) to ensure that distance from the block was one stratum that was considered when determining villages in which to randomly interview children, caregivers, and community leaders. An additional issue was that, due to the lack of pretest quantitative data for some indicators, it was not possible to examine trends in outcomes, such as protection and trafficking, from baseline to endline. One evaluator capacity issue was that

different external evaluators were hired for the baseline and endline evaluation activities, respectively, which negatively impacted the execution of study methods by creating challenges in coordination, standardization, and quality for both data collection tools and methods.

As a context-related issue, it was noted by ChildFund India staff that government officials were reluctant to participate in the consultations for in-depth interviews at baseline. ChildFund India staff asserted that this hesitation was likely due to government officials, primarily those representing the justice and social welfare sectors, fearing reputational risks if they acknowledged and discussed child protection and trafficking issues in areas under their jurisdiction compounded with the belief that it was their "territory" and not the role of civil society to complement child protection awareness and intervention response efforts at the time. This observed reluctance may have impacted data quality and thus not presented a true picture of the protection challenges and risks faced by children, nor an accurate understanding of the context and intertwining interventions and stakeholders responsible for child protection. The evaluation study was also challenged by hiring enumerators (data collectors) who were not fluent in respondents' local dialects and protocols used were not translated into the local language. While respondents had some knowledge of Hindi, ChildFund India's LIPs had to translate questions and responses from participants which required additional time and may have caused issues with fidelity of questions being asked across communities; limited the ability of researchers to adequately probe and/or potentially introduced bias or inaccuracy in responses recorded.

A key organizational lesson learned from the implementation of this community-based study, and relevant for many other ChildFund evaluation studies, was the importance of investing in contracting an external (third party) evaluator, and more importantly, the same external evaluator, for the evaluation study as whole (versus different evaluators for baseline and endline data collection and analysis, respectively), commencing pre-baseline. The study external evaluator can then support ChildFund in considering, selecting, and planning for more rigorous study designs including RCTs—and can provide continuity and standardization in methodological approaches for the duration of the evaluation study. In addition, ChildFund India observed that hiring an independent external evaluator from a different geographic area (who has familiarity with the local dialect) or evaluation team (versus using a local independent evaluator), added an element of credibility to the evaluation, which contributed to building community relationships in support of the evaluation study and to strengthening relationships with local government and community partners beyond the life of the project.

Another lesson was that it is critical to invest enough time and financial resources for onboarding external evaluators, in order for the researchers to adequately understand ChildFund's operating context and to enable the

conduct of more rigorous research methods. Due to the dynamic context of the research setting, external evaluators often need more time during the inception period of the project and its evaluation study to understand the operating context for the study; this has implications for the selection, articulation, and implementation of a study design.

A strength of this evaluation study, despite its one-group pretest posttest design, was the participatory nature of the study and data collection used to better understand and capture contextual issues. Working at the community level to support the execution of the study, ChildFund's LIPs successfully supported evaluation activities from the baseline to endline of this project due to their long-term presence in communities and trust with community stakeholders. Leveraging these relationships was identified as a success factor when introducing external researchers to communities for conducting research activities. It was an important first step in Child-Fund India and LIPs' abilities to garner evidence of community-based child protection interventions to better identify and scale-up such promising interventions.

Reflections on Measuring Impact in Dynamic Contexts and Applying Approaches From This Special Issue

As shown across our three case examples, the community contexts in which we deliver and evaluate the impact of our programming play a critical role in shaping the design, conduct, and overall quality and rigor of the evaluation study. In focusing on approaches for and issues related to conducting RCTs in clinical and community settings, several of the papers in this special issue speak directly to some of the key issues ChildFund has faced and/or needs to pay increased attention to. Moving forward, as we seek to collect evidence on the contribution and attribution of our service programming and expand our organizational evidence-base, we will use this evidence to inform our future program design and delivery and share our knowledge of what "works" and "does not work" externally. The issues featured in this special issue that we intend to pay heightened attention to are: addressing validity threats, identifying alternative study designs, and placing an increased emphasis on process evaluation.

Addressing Validity Threats. An important evaluation design issue that we face overall and have shared across our case examples, particularly, the one-group pretest posttest studies conducted for responsive parenting project in Kenya and Zambia and the child trafficking project in India, is how to strengthen our evaluation designs in order to address validity threats and produce evidence with enough confidence to support causal conclusions and scalability to broader contexts. Wadwha and Cook (this issue) offer a framework for examining threats to validity in RCTs focused on the study of child and adolescent development. Their framework highlights the importance of looking beyond the preservation of just internal validity and

attending to an additional suite of validity threats that include statistical conclusion validity, construct validity, and external validity.

Specific assumptions underlying each of these constructs are particularly relevant for an organization like ChildFund as we attempt to implement the most rigorous studies possible in specific, dynamic contexts and with our pragmatic study design constraints. Thus, this validity framework and its twenty-one assumptions are relevant for us to consider in the conduct of design types other than RCTs. Some of the assumptions that are most relevant to our evaluation work are: (a) for internal validity, minimizing observed differences between intervention and control/comparison groups at baseline (pretest) and minimizing differential attrition; (b) for statistical conclusion validity, selection of correct statistical tests and hypothesis testing at the correct alpha level (which are methodological challenges for some of our evaluation studies due to external evaluator capacity issues); (c) for construct validity, the choice of a control group, particularly one that is not receiving similar services at some dosage level (as, e.g., in our PUENTES Pilot Study); and (d) for external validity, our study population directly maps to and represents the planned targets of the intervention(s).

Identifying Alternative Study Designs. As we consider validity threats and develop a research program that supports the conduct of increasingly rigorous research, we realize that we must consider alternative study designs—both at the higher study design rigor levels, for example, the "gold standard" RCT, as these are often not feasible programmatically and/or financially and the lower study design rigor levels, for example, the one-group pretest-posttest study (as these studies are often most feasible but preclude examining project attribution). For example, an evaluation of a ChildFund preschool-based DRR project in Indonesia (see Proulx & Aboud, 2019) successfully used a quasi-experimental study design that included a posttest-only comparison of a randomly selected group of children for the preliminary study of a dynamic context intervention. A dynamic feature of this project's context is that the island of Sumba, where the preschool intervention was conducted, is frequently affected by disasters such as floods, earthquakes, and landslides. The evaluation study was effectively executed in eight intervention schools and six comparison schools in this setting. Potential confounding factors (e.g., maternal level of education, family assets, child nutritional status) were controlled using regression analyses. The use of this quasi-experimental design also informs and serves as a "stepping stone" for the use a more rigorous study designs in further ChildFund research on this DRR education intervention. Two papers in this special issue share additional options for alternative designs that are extremely relevant for our work in dynamic context settings.

First, Tavecchio et al. (this issue) share, from their work on examining the effectiveness of forensic youth care, the use of Participatory Peer Research (PPR) as an alternative study design option. In PPR, the participant themselves conduct the study to facilitate a quick translation

of the study results into clinical practice. Strengths of this approach are that it yields rapid results (e.g., within a few months) and that it provides participants with a "tool to change their social environment and life circumstances." Thus, the PPR approach may be a good fit for evaluating ChildFund's programming, particularly for initiatives that focus on enabling participants (e.g., adolescents and youth in our Combatting Child Trafficking Project study in India) to be agents of change in their community.

Second, Maric, Geuke, Miočević, Wolters, and de Haan (this issue) share the application of Single-Case Experimental Designs (SCEDs), with the use of innovative data analytic methods, as an alternative to RCTs and other rigorous multiple case studies. Maric and colleagues' paper not only shares the benefits and use of SCEDs but also shows how a set of data analytic methods can be used to identify effective intervention components through the identification of mediators—variables that are linked to change in core intervention outcomes for specific participant cases. In SCEDs, a single participant or just a few participants are tested over a given duration in the study. This study design (and associated data analytic methods) can potentially be used by ChildFund as an approach to pilot and test the feasibility of new programming and innovations before conducting a more resource intensive, larger scale evaluation study, such as a RCT. In fact, we have recently used a qualitative case study approach to examine the commercial sexual exploitation of children in the Philippines (ChildFund, 2018b). Looking forward, the SCED approach may be a way to examine, in a controlled manner, the impact of community-based programming on specific target groups, for example, the effect of our child trafficking project activities on female adolescents in Southern Rajasthan (see our third Case Example).

Placing Increased Emphasis on Process Evaluation.   Many projects aim to better understand "what works" while failing to adequately understand the connections between the intended program design and its effects, which could be addressed through a rigorous process evaluation instead of or in addition to an impact evaluation (Bell & Peck, 2016; Connolly, Keenan, & Urbanska, 2018; Epstein & Klerman, 2016; Goodman et al., 2018). As shared across our three case examples, we had key lessons learned about the importance of process evaluation as a critical component of our evaluation studies, regardless of their level of study design rigor. Process evaluations have been determined as an indispensable "protection against program failure" (Epstein & Klerman, 2016, p. 40), however, in our experience, they are not frequently enough applied to the evaluation study process itself, to protect against the evaluation study failing to be well-conducted.

Ponguta et al. (this issue) share their experience with—and the importance of—including a rigorous process evaluation as a component of a RCT. Their learning from this experience complements and supplements many of our evaluation study "lessons learned" regarding the need for enhanced

implementation research as their process evaluation focused on an intervention (ECD parenting project in a dynamic context—Palestinian refugee and other marginalized communities in Beirut) that are similar to Child-Fund's programming contexts. Their implementation and process evaluation framework explores some of the domains that have been most challenging to us (and where we have identified the need for strengthening within and across our evaluation studies), namely: exploration of context, highlighting, as we have learned, the importance of working closely with community leaders to build trust and enhance participation; quality of project implementation, fidelity, and attendance, identifying the importance of monitoring for "ad hoc" changes to intervention delivery (an issue some or our projects have fallen prey to); and enablers and barriers to the project's evaluation, with an important facilitator being enabling the communities' co-construction of evaluation process and activities, such as the development of an assessment battery.

Moving Forward With Conducting Impact Study in Dynamic Contexts. As ChildFund continues to assess programming impact and moves forward in building our evidence-base, we embrace additional rigor-enhancement strategies such as the increased and proper use of pilot studies (Westlund & Stuart, 2017); the utilization of study designs that have been effectively used for community-based interventions such as the interrupted time series design which enables multiple communities to be repeatedly assessed with an intervention being introduced in one community at a time (Biglan, Ary, & Wagenaar, 2015; Henry, Tolan, Gorman-Smith, & Schoeny, 2017; Shadish, Cook, & Campbell, 2002); increased application of impact study decision trees; and higher standards and more rigorous processes for identifying and selecting external evaluators. We have also placed an increased focus on the use of a tiered evidence model (Haskins & Margolis, 2014; Zandniapour & Deterding, 2018) to grade our progression in identifying program impact and building a cumulative body of evidence and learning—and inform the development and execution of a cross-organization research agenda (ChildFund, 2018a).

Social service project implementation has proven most challenging among vulnerable groups when the intent is to improve outcomes across the social sectors such as health, child development, and education (Mead, 2016). The dynamism characterizing these contexts require projects, and concomitantly their impact evaluation studies, to address participants "where they are" by responding to individual participants' complex situations as well as factors such as "identity, resources, culture, strengths and community" (Goodman et al., 2018, p. 61). In such contexts, participants may receive different services based on their vulnerability classification at the onset of an intervention (ChildFund, 2018c) and/or may access government mandated services or communication campaigns, which makes it more challenging to implement a study design like an RCT. Thus, design and approach alternatives—as well as ways to enhance the

implementation of RCTs, and other study designs, in dynamic contexts (e.g., with the inclusion of rigorous process evaluation) are needed. There are alternative designs to the traditional RCT for assessing impact that have the potential to maintain a high level of rigor and yield valid research in these types of dynamic community settings. They include regression discontinuity designs (RDD) that create a stronger comparison group (and reduce threats to internal validity) in a quasi-experimental design evaluation study by forming intervention and control groups using a well-defined cutoff score (the group below the cutoff score receives the intervention and the group above does not, or vice versa), the previously mentioned interrupted time series design, and case study designs such as the SCED discussed in this special issue.

To conclude, the papers in this special issue provide first-hand and pragmatic examples of approaches to enhance the rigor of evaluations that can be applied by child and adolescent service organizations such as Child-Fund in real-life settings. Identifying alternatives to and facilitators of RCTs is not merely an "academic exercise," instead it is paramount for determining and prioritizing "appropriate methods to uncover evidence that is valid, reliable and meaningful" (Goodman et al., 2018, p. 59) to human service providers. It is equally important to find ways to do valid research in real-life, dynamic settings. For organizations like ChildFund working to strengthen supportive environments and improve outcomes for children, adolescents, and youth in dynamic contexts, this entails an intentional balance of the "ideal" or optimal evaluation design with practical considerations and needs of multiple stakeholders, including, and perhaps most importantly, community participants—and the children, themselves.

References

Bell, S. H., & Peck, L. R. (2016). On the "How" of social experiments: Experimental designs for getting inside the black box. *New Directions for Evaluation*, *152*, 97–107. https://doi.org/10.1002/ev.20210

Biglan, A., Ary, D., & Wagenaar, A. C. (2015). The value of interrupted time-series experiments for community intervention research. *Prevention Science*, *1*(1), 31–49. https://doi.org/10.1023/a:1010024016308

ChildFund International. (2017). *Puentes project: Final report*. Washington, D.C.: Author.

ChildFund International. (2018a). *2018 impact report*. Washington, D.C.: Author.

ChildFund International. (2018b). *Children's journey into the world of commercial sexual exploitation*. Washington, D.C.: Author.

ChildFund International. (2018c). *Endline evaluation report: Assuring the essentials of optimal development for children affected by HIV and AIDS in Kenya and Zambia*. Washington, D.C.: Author.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomized controlled trials in education research 1980–2016. *Educational Research*, *60*(3), 276–291.

Deterding, N. M., & Solmeyer, A. R. (2018). Building evidence in challenging contexts: Introduction to the special section. *American Journal of Evaluation*, *39*(1), 24–26.

Epstein, D., & Klerman, J. A. (2016). On the "When" of social experiments: The tension between program refinement and abandonment. *New Directions for Evaluation*, *152*, 33–45. https://doi.org/10.1002/ev.20213

Goodman, L. A., Epstein, D., & Sullivan, C. M. (2018). Beyond the RCT: Integrating rigor and relevance to evaluate the outcomes of domestic violence programs. *American Journal of Evaluation*, *39*(1), 58–70.

Haskins, R., & Margolis, G. (2014). *Show me the evidence*. Washington, D.C.: Brookings Institution Press.

Henry, D., Tolan, P., Gorman-Smith, D., & Schoeny, M. (2017). Alternatives to randomized control trial designs for community-based prevention evaluation. *Prevention Science*, *18*(6), 671–680. https://doi.org/10.1007/s11121-016-0706-8

Mead, L. M. (2016). On the "how" of social experiments: Using implementation research to get inside the black box. In L. R. Peck (Ed.), *Social Experiments in Practice: The What, Why, When, Where and How of Experimental Design and Analysis. New Directions for Evaluation*, *152*, 73–84.

Peck, L. R., & Goldstein, N. (2016). Social experiments in practice: Introduction, framing, and context. In L. R. Peck (Ed.), *Social Experiments in Practice: The What, Why, When, Where and How of Experimental Design and Analysis. New Directions for Evaluation*, *152*, 9–17.

Proulx, K., & Aboud, F. (2019). Disaster risk reduction in early childhood education: Effects on preschool quality and child outcomes. *International Journal of Educational Development*, *66*, 1–7.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Westlund, E., & Stuart, E. A. (2017). The nonuse, misuse, and proper use of pilot studies in experimental evaluation research. *American Journal of Evaluation*, *38*(2), 246–261.

Wynn, B. O., Dutta, A., & Nelson, M. I. (2005). *Challenges in program evaluation of health intervention in developing countries*. Retrieved from https://www.rand.org/content/dam/rand/pubs/monographs/2005/RAND_MG402.pdf

Zandniapour, L., & Deterding, N. M. (2018). Lessons from the social innovation fund: Supporting evaluations to assess program effectiveness and build a body of research. *American Journal of Evaluation*, *39*(1), 27–41.

DARCY L. STROUSE, PhD, is a senior research advisor at ChildFund International, where she leads the design and implementation of a global research program to support organizational thought leadership and learning on child protection and child development in vulnerable settings. Her primary research interests focus on youth mental and physical health promotion and the role of supportive family and peer relationships in child and youth development and protection.

KATHRYN MOORE, MA, is an international early childhood development (ECD) and education professional working at the nexus of humanitarian and development initiatives. She has provided technical leadership to ECD initiatives in more than 20 countries. Her most recent research focuses on analyzing early childhood education systems in contexts with varying governance arrangements and the extent to which early childhood education is implemented in contexts affected by crises.