

## Special Issue Article

# Children's learning and development in conflict- and crisis-affected countries: Building a science for action

J. Lawrence Aber , Carly Tubbs Dolan, Ha Yeon Kim and Lindsay Brown

Institute of Human Development and Social Change (IHDS), New York University, New York, NY, USA

### Abstract

This paper critically reviews the opportunities and challenges in designing and conducting actionable research on the learning and development of children in conflict- and crisis-affected countries. We approached our review through two perspectives championed by Edward Zigler: (a) child development and social policy and (b) developmental psychopathology in context. The aim of the work was to answer the following questions: What works to enhance children's learning and development in such contexts? By what mechanisms? For whom? Under what conditions? How do experiences and conditions of crisis affect the basic processes of children's typical development? The review is based on a research–practice partnership started in the Democratic Republic of the Congo in 2010 and expanded to research in Niger and Lebanon in 2016. The focus of the research is on the impact of Healing Classrooms (a set of classroom practices) and Healing Classrooms Plus (an additional set of targeted social and emotional learning activities), developed by the International Rescue Committee, on children's academic outcomes and social and emotional learning. We sought to extract lessons from this decade of research for building a global developmental science for action. Special attention is paid to the importance of research–practice partnerships, conceptual frameworks, measurement and methodology. We conclude by highlighting several essential features of a global developmental science for action.

**Keywords:** global science for action, refugee education, research–practice partnership, social–emotional learning

(Received 22 September 2020; accepted 26 September 2020)

### Introduction

Throughout the world, over 500 million children – nearly one child in four – live in areas affected by armed conflict, crisis, and/or disaster (UNICEF, 2016). Globally, half of the children who are out of school live in conflict-affected countries (UNESCO, 2013). Among children in conflict-affected contexts lucky enough to be in school, their learning can be further derailed by the weakened education system, security concerns, and resource constraints (World Bank, 2018). For decades, the humanitarian aid community has focused most of its efforts on providing food, shelter, WASH (water–sanitation–hygiene), and medical supplies to these children and their families, and for good reasons: these are the basic elements required for children's very survival. However, as a result, a scant 2.7% of humanitarian aid was invested in education programming in 2016 (UNESCO, 2017).

While education may seem less urgent or essential than providing for children's basic health and material needs, we argue that it is not. Besides basic survival, children and their families need to see real, concrete pathways beyond survival to a decent and meaningful future. Formal and informal education can provide pathways to

academic learning and social–emotional development, which in turn can provide pathways to positive development in youth and adulthood. Academic learning and social and emotional development are the twin missions of childhood “beyond survival” in the 21st century. Generally, when children who live in communities and countries facing emergencies (e.g., internally displaced and refugee children fleeing armed conflicts, or children caught in the midst of pandemics and economic shocks) but are nonetheless *in school*,<sup>1</sup> they are better off than out-of-school children (Burde & Linden, 2013; Winthrop & Kirk, 2008). They are in relatively safe/supportive and predictable/cooperative environments designed (however imperfectly) for their learning and development; they have the potential for their significant social and emotional development and mental health needs to be addressed; and they have the opportunity to develop the literacy and numeracy skills needed for a viable future after conflicts and protracted crises recede. In short, they have more opportunities to pursue the core missions of childhood and to experience a more tolerable present and a more promising future.

In the last three decades, (a) under the normative influence of both the Convention on the Rights of the Child (CRC) (articles 28 and 29 on the rights to education (UNCRC, 1989)) and the Sustainable Development Goals on education (SDG4), which pledges to provide all children a quality primary and secondary

<sup>1</sup>Of course, the very concept of being *in school* is now being reinvented throughout the world in the midst of the Covid19 epidemic.

**Author for Correspondence:** J. Lawrence Aber, 246 Greene Street, 417E, New York, NY, USA; E-mail: [la39@nyu.edu](mailto:la39@nyu.edu)

**Cite this article:** Aber JL, Tubbs Dolan C, Kim HY, Brown L (2021). Children's learning and development in conflict- and crisis-affected countries: Building a science for action. *Development and Psychopathology* 33, 506–521. <https://doi.org/10.1017/S0954579420001789>

education (UNDESA, 2015), and (b) as the numbers of countries (and hence the number of children) affected by conflicts and crises have grown, so too have efforts devoted to promoting children's academic learning and their holistic development through access to high-quality education. In this paper, we review the opportunities and challenges we and our collaborators have faced over the last decade in designing, conducting, and using research on how to enhance the learning and development of children in crisis-affected countries via educational programming. We also extract lessons learned from this work for future efforts.

We approached these tasks guided by two interrelated perspectives pioneered and championed by Edward Zigler.

First and foremost, we were guided by Zigler's promotion of the subdiscipline of child development and social policy (Aber, Bishop-Josef, Jones, McLearn, & Phillips, 2007; Zigler & Hall, 2000): Zigler encouraged developmental scientists to fully engage with policymakers and practitioners from various sectors and scholars from various disciplines to better understand program and policy issues that affect the learning, development, and well-being of children. As one shining example, Zigler and colleagues drew on the best of the developmental, health, and education sciences and practices at the time to inform the design, evaluation, and improvement of the federal Head Start program for young, low-income children (Zigler & Styfco, 2010) and used the results of evaluations of Head Start to propose needed expansions of and important revisions to its design. The child development and social policy perspective poses many important questions about children in conflict- and crisis-affected countries. Here, we focus on lessons from our own impact evaluations and implementation studies of novel education programming for children affected by conflicts and crises. The classic questions of the prevention and intervention sciences are critical here.

What education interventions work to improve the learning and development of children in conflict- and crisis-affected countries (impact)?

How do they work (by what mechanisms)?

For whom (heterogeneity of impact)?

Under what conditions (contextual effects)?

Are such interventions scalable (generalizability)?

In this paper, we explore such questions by reporting selected findings and lessons from an ongoing program of research on the impact of education interventions designed to improve both academic and social-emotional learning (SEL) of children affected by armed conflicts and protracted crises. This research began in the Democratic Republic of the Congo (DRC) in 2011 and then extended into a second phase of research in Lebanon, Niger, and Sierra Leone in 2016.

Second, we were informed by the perspective of developmental psychopathology in context, which powerfully informed Zigler's own research (Luthar, Burack, Cicchetti, & Weisz, 1997; Zigler & Glick, 2001). Stimulated by ecological theorists like Bronfenbrenner and Lerner, this perspective was later extended to incorporate *contexts of development* by Cicchetti and others in the 1990s (Cicchetti & Aber, 1998). It has contributed to our evolving understanding of the development of children exposed to conflicts and crises (Cummings, Goeke-Morey, Merrilees, Taylor, & Shirlow, 2014; Masten & Narayan, 2012). Many salient questions on child development in crisis settings are raised by the developmental psychopathology in context perspective and include the following.

How do experiences and conditions of crisis (e.g., living as a refugee or internally displaced person in a conflict-affected country; living for extended periods in lockdown due to a deadly viral pandemic) affect the basic processes of children's typical development?

To what extent do children's cultures, communities, and atypical experiences lead to important differences in their developmental processes and trajectories?

When children in conflict- or crisis-affected contexts develop differently, to what extent do such differences confer advantages or disadvantages in their specific developmental contexts? Are the differences "adaptive," "maladaptive," or both?

In this paper, we explore these questions by highlighting findings from our research examining associations among risks faced by Syrian refugee children in Lebanon, their developmental processes, and academic outcomes (Kim, Brown, Tubbs Dolan, Sheridan, & Aber, 2020). While we found some similarities with educational and child developmental research in western contexts, we also identified areas of divergence. Although further research is needed to unpack and interpret these findings, we believe they are critical to help guide the design of more impactful and relevant education strategies for children in crisis contexts.

Both the perspectives of child development and social policy and developmental psychopathology in context are fundamental to building a science for action on behalf of children affected by armed conflicts and protracted crises. However, in our opinion, they are not the only building blocks. As we tell the story of our efforts to develop and expand a program of research on primary school-aged children's development in crisis contexts, we will highlight four themes that we have found are critical for ensuring science is transformed into action.

1. Long-term research-practice-policy-funder partnerships – and consortia of such partnerships – are an effective strategy for increasing the generation and use of high-quality evidence for program and policy decision making.
2. Conceptual frameworks and research agendas for the field are necessary for organizing, coordinating, and prioritizing strategic accumulation of research that can inform action.
3. Context-relevant measures and methodologies are required to make progress on child development and education research agendas in low-resourced, crisis-affected contexts.
4. Stakeholders within partnerships must build in time for joint reflection of the research findings and the partnership itself.

## Background

When we embarked on this line of research, what was the state of the knowledge base on what works to promote children's learning and development in armed conflicts and protracted crises? The short answer is too little for research to effectively guide action (Tubbs Dolan, 2018). As noted by Masten and Narayan (2012), in 2012 there was an acute shortage of high-quality evidence on interventions for children exposed to conflicts and crises, especially given the scope of the conflicts and crises throughout the world and the scale of humanitarian efforts to intervene. The majority of rigorously evaluated interventions for children in conflict-affected contexts focused on improving children's mental health and psychosocial wellbeing (Betancourt & Williams, 2008; Jordans, Tol, Komproue, & Jong, 2009). There were fewer than 15 Randomized-Control Trials of interventions that sought to

improve access to and the quality of education for children in conflict-affected countries (McEwan, 2015; Torrente, Alimchandani, & Aber, 2015). Of note, there were no studies at the time that examined the impacts of interventions on both academic learning and SEL – the twin missions of childhood. This was in striking contrast to the large and high-quality evidence base of rigorously evaluated interventions to promote both academic learning and SEL in the USA and other western and high-income countries (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Wigelsworth *et al.*, 2016). Jordans *et al.* (2009) echoed the calls for more rigorous research on the efficacy and effectiveness of interventions in conflict and crisis contexts (what works?) but also noted with alarm “the complete lack of mechanisms research” (p. 10). Mechanisms research – rigorous tests of whether an intervention changes the mediating processes by which the intervention is theorized to causally influence the outcomes – is critical both to a full scientific understanding of the efficacy and effectiveness of an intervention and to guide successful replications of intervention strategies.

This state of affairs did not escape the notice of major bilateral and multilateral funders of humanitarian efforts on behalf of children exposed to conflicts and crises. For instance, in 2013, the United States Agency for International Development (Olenik, Fawcett, & Boyson, 2013) argued that new and better research was needed – research using rigorous experimental or quasi-experimental designs, better measures of key constructs and longitudinal follow-ups – to create a valuable evidence base of knowledge for action. As the value and feasibility of conducting randomized field experiments of education programming to improve the learning and development of children in armed conflicts and protracted crises became more tangible, the lack of valid, reliable, linguistically and culturally adapted and feasible measures of key mediating processes and outcomes became more apparent (Tubbs Dolan & Cairns, 2020).

This raises the question of why we didn't know more about what works, how it works, for whom, and under what circumstances. The answer is primarily because this is difficult work. Even if researchers are resolved to conduct rigorous, well-instrumented, longitudinal impact evaluations of what works to improve children's learning and development and how, they face significant challenges in doing so. Logistically, the recruitment and ethical consent of children and schools in areas of conflicts and crises, the travel time to get to remote locations, the lack of a telecommunications infrastructure to provide any kind of electronic support for data collection, and the need to be smart and ethical in protecting children, teachers, schools, and research staff from danger make research in conflict and crisis contexts very difficult. Financially, such research entails added expenses as well as increased risk – a difficult combination for many donors to tolerate and adequately fund. Such research also raises challenges from political and governance perspectives. Will governments and/or opposing militia permit such research? Who in the family, school, community, or government has the right and the legitimacy to approve or disapprove the conduct of the research? Methodologically, how can researchers, their community, and practice partners ensure that sound designs, procedures, and analyses are practiced under extreme conditions? How do researchers engage all concerned parties, including the children themselves, with learning about and problem solving on the many issues raised by ethically sound and scientifically rigorous research, especially in conflict and crisis contexts? Last but not least, how do researchers and practitioners surmount such

logistical, financial, political, and methodological challenges to research in humanitarian contexts that do not typically value or see a role for such research?

Fortunately, some international non-governmental organizations (NGOs), including one of our principal practice partners, the International Rescue Committee (IRC), have identified these challenges and have begun to mount ambitious but nonetheless partial responses to them. Of note, the IRC created a research, evaluation, and learning unit (now called the Airbel Impact Lab) within the organization that both conducted its own research and created and managed research–practice partnerships with independent university-based researchers. Later, the IRC publicly committed to using evidence on what interventions work in humanitarian situations where it exists and generating evidence where it doesn't exist in all five domains of IRC programming – education, health, economic wellbeing, empowerment, and safety. Currently the IRC works in over 40 countries around the world affected by armed conflicts and protracted crises.

In broad strokes, this was the state of this nascent field of research a decade ago. With this background on the state of the knowledge base at the time on what works to promote children's learning and development in conflict and crisis contexts and on the significant challenges in designing and conducting rigorous research in such contexts, we now turn to our first major collaborative study with the IRC to help expand this knowledge base.

## Healing Classrooms (HC) in the DRC

### *The initiative*

Well before researchers began to study what works in educational programming for children in conflicts and crises, local and national governments and local and international NGOs were devising strategies they believed could be effective. Based on years of qualitative action research in four conflict-affected countries, the IRC developed a model for education in armed conflicts and protracted crises that they came to name “Healing Classrooms” (HC). HC was premised on the notion that, for children affected by armed conflicts and protracted crises, psychosocial and mental health needs must be effectively addressed if they were going to be free to learn academically. The two major components of HC were (a) high-quality reading and math curricula infused with classroom practices that also promote SEL and (b) in-service teacher professional development supports for the reading and math curricula and SEL classroom practices. In addition, in the HC approach, the IRC facilitated ongoing peer-to-peer support for teachers in implementing the curricula and practices via teacher learning circles (for a fuller description of HC, see Aber, Torrente, *et al.* (2017) and Torrente, Johnston, *et al.* (2015)). Like SEL programs in the west, the IRC reasoned that HC would promote both greater academic learning and SEL. While their qualitative action research suggested that HC was a promising strategy to do so, the IRC wished to put HC to a more rigorous test of efficacy and/or effectiveness. Was the HC approach more effective than business-as-usual education programming in conflict- and crisis-affected countries? An opportunity soon arose to conduct such a test.

In 2010, a request for proposal was released by USAID to improve the quality of primary education in about 350 schools expected to serve 500,000 children over a 5-year period in the eastern DRC. The eastern DRC had been at the epicenter of several decades of armed conflict and protracted crises in what has

been referred to alternatively as the Great Lakes War and Africa's World War. Across nine nations, local warlords, competing militias, and national armies fought over rare and valuable minerals, natural resources, and physical control over high-value territories (Prunier, 2009). By 2010, two decades of conflict – which had substantially damaged the physical and human infrastructures of the schools in the affected communities across these countries – began to subside enough to make rebuilding the education infrastructure desirable and viable. The IRC's education unit thus proposed implementing HC in DRC and IRC's research, evaluation, and learning unit proposed to researchers at New York University (NYU) that we collaborate in conducting a rigorous impact evaluation of HC if the project were to be funded by USAID. We readily accepted the proposal.

Thus began a research–practice partnership that, over the next 4 years, would design, implement, and begin to analyze results from a large and complex field experiment of the HC approach to primary education in the eastern DRC. The results have been reported in phases in a series of policy briefs and peer-reviewed journal articles from 2015 to 2018. In the remainder of this section, we summarize the intervention theory of change (TOC), the evaluation design, and key findings before discussing lessons learned and next steps.

### The TOC

In order to evaluate the impact of HC on children's academic learning and SEL, the IRC and NYU researchers needed to co-construct a reduced form of the TOC. Because this evaluation was both large in scale and poorly funded, we could not measure all elements of a comprehensive TOC. Rather, the practice and research staff of IRC needed to specify the core or essential elements of the TOC to guide the many decisions that the IRC had to make to implement HC in 350 schools over a large geographic region; the research partners from NYU had to create a research design that was both scientifically rigorous and practically feasible and to develop or adapt measures of the core elements of the TOC. In short, the program design and the research design needed to be iteratively co-constructed to meet both the intervention and the evaluation requirements. The reduced-form TOC that emerged from our joint planning process is presented in Figure 1. As can be seen, the theory hypothesized that HC would positively impact two key elements of children's school ecologies: (a) children's perceptions of their school environments (here including their perceptions of schools as safe/supportive and predictable/cooperative) and (b) teachers' motivation and wellbeing. In turn, changes in these two critical elements of the school ecology were hypothesized to lead to impacts on children's academic skills (reading and math) and social–emotional wellbeing (here operationalized as mental health symptoms and experiences of victimization). Importantly, the TOC specified that not all of the impact of HC on children's academic learning and social–emotional functioning would be mediated by the children's perceptions of their school ecologies and their teachers' motivation and wellbeing. Undoubtedly, there were other mediators we couldn't measure or couldn't even yet hypothesize!

### Research design

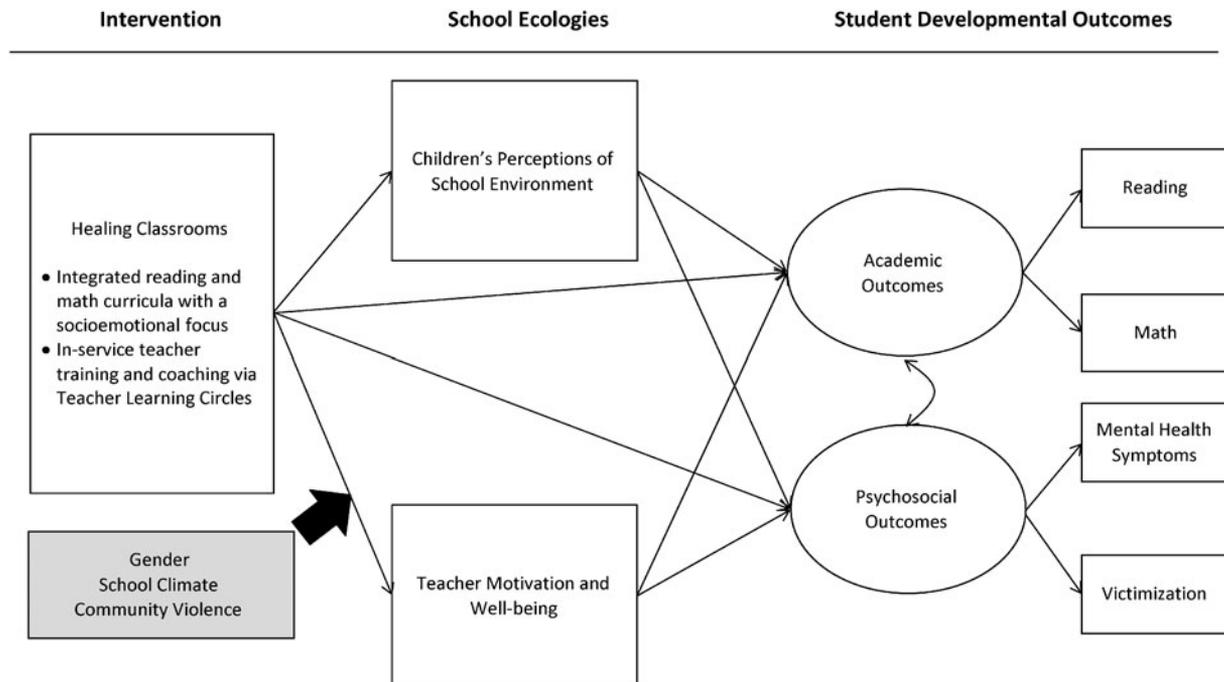
Historically, most humanitarian relief funders and organizations measure their success by the number of beneficiaries they serve (outputs), not by the quality or the impact of the services

they provide (outcomes). In addition, humanitarian relief organizations often look with suspicion at no-treatment control groups and at experimental impact evaluations of their programs and policies because of ethical concerns about withholding potentially beneficial treatments in such trying conditions. These and other factors reduce interest in and support for rigorous field experiments of the impact of service strategies on child and human development outcomes in humanitarian settings. This raises the questions of what happens when there are many more potential beneficiaries than there are resources to serve them or what happens when the implementing organization's institutional capacity for universal implementation for all beneficiaries at the start of an intervention does not exist. Under such circumstances, might a cluster randomized trial with a waitlist control design – in which the intervention is rolled out over time – prove to be both ethically sound and scientifically valuable? Somewhat surprisingly, USAID, the Ministry of Education in the DRC, IRC decision makers at country, regional, and HQ levels, and the NYU all decided – albeit for different reasons – that the answer was yes. Such a field experiment of HC was deemed possible and desirable, but the field experiment needed to be constructed under serious budgetary and logistic constraints.

The full details of the experimental design we used are provided by Torrente, Johnston, et al. (2015), Wolf et al. (2015) and Aber, Torrente, et al. (2017). Here we describe several features of the experimental design to illustrate some of the key challenges in field experiments of novel education programming in conflict- and crisis-affected countries. At the time, for administrative reasons, the targeted 350 schools were already grouped by DRC's Ministry of Education into school administrative clusters of two to six schools based on their geographic proximity. Such clustering would facilitate the logistics of teacher training for quality improvement and school monitoring for accountability. We were eager to deliver HC in all schools in such existing clusters (a) to enhance generalizability of the findings (external validity – we were evaluating HC in the way it would be delivered if it went to scale nationally) and (b) to protect the experiment from potential for contamination/spillovers of the intervention from treatment to control schools in close proximity (internal validity). Thus, school clusters became the policy-relevant programmatic unit of intervention delivery and thus the scientific unit of randomization and analysis.

Very importantly, not all the school districts were judged to be ready to start implementation in the first year: some districts were still too vulnerable to renewed armed conflict and some were not yet sufficiently organized to coordinate with field researchers. Two cohorts of school clusters were thus created. Cohort 1 consisted of clusters in four school districts ready to start in year 1, while Cohort 2 consisted of clusters in four other school districts that would not be ready to start until year 2. Then, through independent public lotteries in each of the participating school districts, clusters of schools were randomly assigned to start implementing HC in one of three academic years: 2011–2012, 2012–2013, or 2013–2014 for Cohort 1 and 2012–2013 or 2013–2014 for Cohort 2. Finally, we decided to collect data on randomly selected children in grades 2–5 at three time points at the end of each academic year.

This ambitious design (arrived at via complex tradeoffs among administrative feasibility, scientific rigor, and budgetary cost) resulted in three unique tests of the impact of HC on children and their teachers: in Cohort 1 (immediate readiness) schools after 1 year of intervention, in Cohort 2 (delayed readiness) schools after 1 year of intervention, and in Cohort 1 (immediate



**Figure 1.** Reduced form of theory of change (TOC) for Healing Classrooms (HC).

readiness) schools after 2 years of intervention. With this design, we were able to pose and partially answer the four classic questions of prevention and intervention research – what works, by what mechanisms, for whom, and under what conditions?

Does HC improve children’s school ecologies, teacher wellbeing, and children’s academic outcomes and social–emotional skills? (what works?)

Is there evidence to suggest that changes in children’s outcomes are due to changes in children’s school ecologies or teacher wellbeing? (by what mechanisms?)

Is there evidence of variability in the impact of HC for different types of children and teachers or in different types of school? (for whom? under what conditions?)

The details of the multivariate, multilevel statistical analyses employed to answer these questions are available in earlier papers (Aber, Torrente, et al., 2017; Aber, Tubbs, et al., 2017; Torrente et al., 2019; Torrente, Johnston, et al., 2015; Wolf et al., 2015). In the next part of this paper, we summarize the answers to these four classic questions we were able to derive and discuss our partnership’s collective interpretation of the results (Table 1).

### Key findings from the evaluation of HC in the DRC

From the child development and social policy perspective and through a deeply collaborative research/practice partnership, we knew more at the end of the evaluation in the DRC than we did before about how to improve children’s academic learning and SEL in conflict and crisis contexts. The results indicated that the program showed considerable promise, but also considerable room for improvement.

#### What works?

After 1 year of implementation in the two different cohorts, HC improved some aspects of children’s perceptions of their school

ecologies. For the designers and implementers of HC (the IRC), these changes were what they considered to be goal #1 in education programming for children affected by armed conflicts and protracted crises. They reasoned that more safe/supportive and predictable/cooperative school ecologies were necessary to free children cognitively and emotionally to grow academically in school.

As hypothesized by the HC reduced-form TOC, in Cohort 1 and again replicated in Cohort 2, children improved in basic literacy skills (by an average of 3–4 months of additional growth per year compared with the rate of learning in the control group in eastern DRC) and numeracy skills (by an average of 9–10 months of growth per year) (Aber, Torrente, et al., 2017; Torrente et al., 2019; Torrente, Johnston, et al., 2015). These are promising improvements under such difficult conditions, but it must be noted that the annual rates of improvement in these skills in control schools in the eastern DRC are very slow indeed. It is thus likely that even if children experienced high-quality HC every year of primary school – and thus achieved this enhanced rate of improvement in basic skills – most children would nonetheless leave primary school functionally illiterate and innumerate.

#### What doesn’t work (yet)?

Neither in Cohort 1 nor in Cohort 2 did HC significantly improve teachers’ job-related wellbeing (motivation, job satisfaction, burn-out) or children’s social–emotional outcomes (mental health symptoms or experiences of peer victimization). It thus appears that, while HC is sufficient to effect positive changes in children’s perceptions of their school ecologies and their basic academic skills, it is not sufficient to move the needle on the social–emotional outcomes for teachers or their students. In retrospect, perhaps this should not be surprising. While HC aimed to improve the quality of the classroom environment, it did not explicitly or directly target the key physiological, cognitive, social, and emotional mechanisms known to predict children’s or teachers’

**Table 1.** Summary of impacts of HC by year.

	Cohort 1 1 Year	Cohort 2 1 Year	Cohort 1 2 Years
School ecologies			
<b>Child perceptions of school environment</b>			
Safe and supportive	+	?	NS
Predictable and cooperative	–	+	NS
<b>Teacher motivation and wellbeing</b>			
Teacher motivation	NS	NS	NS
Teacher burnout	NS	NS	NS
Teacher job satisfaction	NS	NS	NS
Child outcomes			
<b>Academic outcomes</b>			
Literacy skills	+	+	NS
Math skills	+	+	NS
<b>Social and emotional outcomes</b>			
Mental health symptoms	NS	NS	NS
Peer victimization	NS	NS	NS

Note: + indicates positive impact of Healing Classrooms (HC); – indicates negative impact of HC; NS indicates no impact of HC.

mental health and wellbeing. In crisis-affected contexts, where children and their teachers face extreme adversities that can overwhelm fundamental human adaptive systems, a focus on the safety and supportiveness of the classroom environment may be necessary but insufficient to improve children's and teachers' wellbeing without attention to other powerful causal mechanisms (Tubbs Dolan & Weiss-Yagoda, 2017).

When the program was extended for an additional year in Cohort 1, the initial positive impacts of HC on reading, math, and school supportiveness faded out (Torrente et al., 2019). Retrospective qualitative data collection suggested that the quality of implementation probably suffered in the second year due to the addition of new treatment schools from the waitlist condition at the same time as the number of IRC staff dedicated to teacher training and monitoring declined. Unfortunately, we were unable to collect the prospective implementation data that would help evaluate this issue.

#### *By what mechanisms?*

The HC impacts we observed on literacy and numeracy skills after 1 year were plausibly and partially mediated by changes in children's perceptions of their school ecologies (Aber, Tubbs, et al., 2017). We say "plausibly" because with this TOC, research design, and the set of statistical assumptions required, we could not formally test true causal mediation. While our estimates of the impact of HC on change in the mediators and on change in the outcomes are causal, our estimates of the impact of the change in mediators on the change in outcomes are correlational. This is an unfortunate yet unavoidable limitation that can only be addressed via replication and accumulation of research. We say "partially" because there is still an unexplained direct effect of HC on outcomes even after accounting for the indirect effect via children's perceptions of their school ecologies. This means that there are likely other mechanisms that helped children in

HC to learn better, which the current TOC and the data were not able to capture – suggesting the current TOC is incomplete.

#### *For whom and under what conditions?*

The main take-home message is that HC had equally positive effects across important subgroups of children (age, grade, gender) and in two independent cohorts of school districts in the country. However, there were two exceptions to this pattern.

First, when we tested gender differences in impacts of HC on teacher wellbeing, we discovered that the null average impacts were obscuring opposite effects for males and females (Wolf et al., 2015). As hoped, HC increased teacher wellbeing among the male teachers; unexpectedly, HC decreased wellbeing among the female teachers. Although we did not have adequate data to investigate why this may be so, anecdotal evidence from IRC field staff generated a hypothesis. Because the teacher corps and consequently the teacher learning circles (the process for peer coaching and support) were 75% male on average, participation in the teacher learning circles may have been more stressful and less supportive for female teachers who were in the minority.

Second, HC had positive impacts (a) for both minority-language and majority-language children and (b) for schools with higher and lower percentages of minority-language children. Nonetheless, minority-language children and children in schools with higher percentages of minority-language children benefitted more than the majority-language children and schools. It is important to note here that the DRC is an enormously linguistically diverse nation with over 215 living mother tongues and four indigenous national languages. This is one reason why French has been and remains the national language of instruction in the DRC. Thus, among refugee and forcibly displaced children, a significant proportion of children did not speak the same mother tongue as their teachers and were only beginning to learn French.

In conclusion, the results of the impact evaluation were generative. We learned that the positive impacts of HC in two independent cohorts of Congolese schools on children's perceptions of their school ecologies and on their literacy and numeracy skills were consistent with two central elements of the co-constructed TOC that guided the impact evaluation. However, the findings about the lack of impact on teachers' and children's social-emotional wellbeing and the fade-out effects after 1 year were not consistent with the TOC. Moreover, despite our adherence to the child development and social policy approach, we lacked the ability to adequately explain *why* and *what to do about it* – key impediments to developing a science for action. We now turn to our reflections on what succeeded and what we needed to improve as we continued on our path towards a science for action.

#### *Lessons learned towards a science for action: HC in the DRC*

As with the research findings themselves, we also learned a lot from the evaluation research in the DRC about how to build a science for action for children in conflict- and crisis-affected contexts. We reflect on these lessons in the context of the four themes we highlighted at the start of the article.

#### *Partnerships*

The results of the evaluation provided an existence proof that research–practice partnerships can be an effective and ethical strategy for generating rigorous evidence at scale under some of the most volatile and dangerous conditions in the world. Humanitarian organizations like the IRC have decades of

experience operating and reaching children in crisis contexts. They are also attuned to the ethical principles underlying the provision of services to and collection of information from children, families, teachers, and schools in such contexts. Researchers bring to bear theories and evidence of how children develop, experience leveraging the realities of limited program and research funding to create rigorous research designs, and expertise developing new or testing existing measurement tools to assess program outcomes reliably and validly. The resulting evidence would not be possible without either partner. However, such partnerships require tremendous perspective taking, given that the imperatives of practitioners and researchers are often at odds. Humanitarian actors must respond to the needs of children in crisis as quickly and as widely as possible, while researchers strive to attain unbiased and precise estimates of intervention impacts – a traditionally long and slow process. Balancing practical feasibility and scientific rigor requires a shared understanding of priorities, a commitment to learning and listening, and a deep respect for respective strengths. In the DRC, we were able to embark on such a relationship with the IRC, but we also recognized two limitations to our partnership model at the time.

First, both the IRC and NYU-TIES (Transforming Intervention Effectiveness and Scale) are organizations grounded in western, educated, industrialized, rich, democratic (WEIRD) traditions and contexts (Henrich, Heine, & Norenzayan, 2010; Wuermli, Tubbs, Petersen, & Aber, 2015). Western researchers and practitioners tend to call non-academic skills “SEL,” but countries from the global south tend to frame the non-academic features of education in quite different ways, such as peace education, moral education, or citizenship education (Torrente, Alimchandani, et al., 2015). In efforts to rapidly intervene and conduct research in crisis contexts, we risk insufficiently grounding programmatic and research approaches in local cultures, frames, priorities, norms, and values while reifying existing inequities in who has access to research and research capacity.

Second, and relatedly, we learned that such research–practice partnerships are only one small part of the larger political economy of countries in conflicts and crises. Other actors such as policymakers and donors are able to enhance or constrain the effectiveness of the partnership and the research (e.g., by cutting key funding part way through the project). This suggested to us that a diversity of partnerships and networks of partnerships – engaging both local and regional actors – may be necessary to move as a field towards a science for action.

### *Conceptual frameworks*

As described above, we co-constructed with the IRC a reduced-form TOC that allowed us to specify the core or essential elements of HC and create a research design that was both scientifically rigorous and practically feasible. While we did the very best we could at the time, in retrospect we did not ground the work sufficiently in theories and methods from developmental psychopathology. The field of developmental psychopathology had learned a lot about risk and protective factors facing children and their families at multiple levels of the human ecology. The field also had learned a lot about the many cognitive, social, and emotional developmental processes that mediate the influence of risk and protective factors on child outcomes, including children’s academic outcomes and social and emotional skills. The inclusion of such risk and protective factors and developmental processes in our TOC and measurement approach could have helped more fully and precisely test the impact of the intervention on children’s

learning outcomes in our mediation models. It also could have provided key insights into the lack of impact on social and emotional outcomes. Our failure to do so was in part a reflection of logistical and budgetary constraints in this particular study, but it is also a reflection of existing inequities in where, with whom, and for whom developmental science research has taken place: 95% of psychological research has occurred with just 5% of the world’s population (Arnett, 2008). Therefore, we could not draw much on prior theoretical, methodological, and empirical research on typical and atypical child development in the global south, including in linguistically and culturally diverse countries affected by armed conflicts and protracted crises, to inform the design and conduct of this important impact evaluation. This is a shortcoming of our study and a shortcoming of the developmental science field at large – it is not yet a global developmental science.

### *Methods and measures*

Regarding research design, methods, and measures, we learned from this evaluation what it actually takes to conduct high-quality research in crisis-affected contexts – and we learned we needed to do significantly better. First and most foundationally, administrative data systems with unique identifiers that enable tracking of children over time were not available in schools in the DRC (or in the majority of crisis-affected contexts due to a confluence of lack of data infrastructure and migration). Investing in such systems under the evaluation budget was not feasible and thus we were unable to reliably and feasibly track individual children over the 3 years of the study. This prevented us from examining individual child-level dynamics nested within dynamic school contexts, one of the central foci of the developmental psychopathology in context perspective. Second, due to budgetary and logistical constraints, we were only able to collect annual waves of data on representative samples of children in second to fifth grade. This meant we could not evaluate within-year change – only between-year change – thereby obscuring key fluctuations in outcomes within a school year that could shed light on the ultimate positive or null impact findings.

Another serious shortcoming was our inability to conceptualize and confidently measure children’s holistic learning and development (CHILD) outcomes and the quantity and quality of implementation of HC. The former concern stems, in part, from our use of CHILD measures such as the Early Grade Reading Assessment (Gove & Wetterberg, 2011) and the Strengths and Difficulties Questionnaire (Goodman, 1997) – measures that were developed for use based on, respectively, theories of early literacy development and social–emotional wellbeing in western contexts. We elected to use these measures (which had been used previously in some low-income and conflict-affected countries) given the lack of existing validated measures in the Congolese context and the need to rapidly mount the impact evaluation with inadequate funding for prospective measure development. While the measures were adapted for use in the Congolese contexts, the resulting data did not meet the assumptions (e.g., normal distributions) necessary for traditional factor analytic and scoring approaches (in the case of the Early Grade Reading Assessment, Halpin & Torrente, 2014) and did not adhere to previously hypothesized factor structures (in the case of the Strengths and Difficulties Questionnaire, Torrente, Johnston, et al., 2015). To address such concerns, we applied post-hoc state-of-the-art psychometric methods that enabled us to build confidence in the reliability and validity of the data. Nonetheless, it is possible

that we did not see an impact of the program on social and emotional wellbeing, for example, because we did not measure it in a contextually or developmentally valid way. In addition, we were unable to collect any information on the dosage and quality of implementation of HC. Thus, we don't know (a) whether variation in the size of impact of HC is explained in part by variation in its implementation or (b) whether the fade out of positive effects after 2 years of the HC intervention was due to a decline in the quantity or quality of implementation. Such experiences clearly indicated that more investment was needed to develop and adapt field-feasible measures of program implementation and quality and CHILD with evidence of reliability and validity in crisis contexts (Tubbs Dolan, 2018).

#### *Joint reflection*

Traditional models of partnership are often transactional and communications one-sided: funders and/or practitioners commission an evaluation, and researchers report results at the end of project for “take-up.” For both tangible and intangible reasons, our partnership with the IRC evolved as one that strove to be more deeply collaborative. Our joint reflection on both the positive and negative results from the DRC work and the strengths and shortcomings of the partnership showed us how much we were able to achieve together – and how much more work can be done as a partnership and as a field. What came next?

### **Education in Emergencies: Evidence for Action (3EA)**

#### *The initiative: overall*

We often liken the evolution of our research–practice partnership with the IRC to a relationship. After “dating” for 4 years we decided to take the next step – to formally commit to a long-term relationship. This relationship was advanced through our next joint strategic initiative, 3EA. Launched in 2016 with start-up funding from Dubai Cares, 3EA was designed to marry high-quality program delivery approaches and perspectives with rigorous research following an iterative process of design, implementation, evaluation, and reflection. Building on all we had learned through our earlier partnership, 3EA encompassed three intertwined areas of research, each of which focused on different contexts using various partnership models.

- 1. Impact of education programming on CHILD in conflict and crisis contexts.** We learned in the DRC evaluation that HC programming was necessary but not sufficient: it needed to be adjusted and tested again in order to ensure impacts on both academic outcomes and social–emotional skills. As such, we experimentally evaluated a set of revised and newly devised education and SEL programs to continue to identify what works (how, for whom, and under what conditions). In order to better combine both the perspectives of child development and social policy and development and psychopathology, we did so using theories of change (TOCs) and research designs that also permitted us to pose and answer questions about children's normative development. We undertook this work with the IRC in two different contexts – with Syrian refugee children in Lebanon and with Nigerien host community and internally displaced children and Nigerian refugee children fled from Boko Haram in Niger.
- 2. Implementation of education programming in conflict and crisis contexts.** We learned from the DRC evaluation that we

needed to better monitor how programs were implemented in order to provide timely course-correction data and to make sound inferences about program impact. To make progress towards this goal, we conducted an intensive implementation study of IRC education programming in Ebola-affected communities in Sierra Leone. This involved clarifying and mapping the planned implementation of key components of the education program, and then collecting daily information about children's and teachers' attendance and classroom activities. In turn, we were able to examine what predicts variation in program implementation, as well as whether and how such variations forecast children's academic and social–emotional outcomes.

- 3. Measurement of program implementation and children's holistic development in conflict and crisis contexts.** We learned from the DRC evaluation that we needed better measures of CHILD outcomes and program implementation for use in crisis contexts. We also learned that a variety of partners needed to be involved in such work. As such, in our partnership with the IRC, we worked intensively to develop, adapt, and test measures in the impact evaluations in Lebanon and Niger. We also convened and funded a consortium of seven other diverse research–practice–policy partnerships working in the Middle East, North Africa, and Turkey to make progress on a field regional measurement agenda. In doing so, we aimed to provide a supply of open-source measures with evidence of cultural sensitivity, reliability, and validity while building the regional research capacity and networks necessary for a science for action.

Each of these three areas of research are complex and important to understand in their own right, while the sum of all three is greater than the parts – and greater than the space limitations of a journal article! As such, we focus for the remainder of the paper on the background, key findings, and lessons learned from our research on the impact of education programming in Lebanon and Niger. In doing so, we highlight the role that program implementation and measurement played in the impact evaluations. Interested readers are referred to fulsome discussions in Tubbs Dolan and Cairns (2020) on the measurement component and in Brown, Kim, Annan, and Aber (2019) and Wu and Brown (2020) on the implementation research component of 3EA.

#### *The initiative: focus on impact evaluations in Lebanon and Niger*

The education programming tested in Lebanon and Niger expanded and built on key findings from the evaluation in the DRC in two main ways.

First, we recognized that the strategy for providing education in crisis contexts has shifted. As host-country governments open up space in their formal school systems for refugee students, NGOs are increasingly engaged in providing or supporting complementary education programs that can help refugee and host-country children navigate barriers to learning and retention in formal schools (DeStefano, Moore, Balwanz, & Hartwell, 2007). However, while governments and other actors have made significant investments in improving access to education through the formal system, little rigorous research has examined how such complementary education programs can support children's learning in formal schools in crisis contexts. Such research is imperative for informing politically- and resource-viable education

strategies. In Lebanon and Niger, the IRC thus worked to adapt their basic HC approach (hereafter called HCB (Healing Classrooms basic)) to be delivered through after-school, remedial education programs. Such work was done in coordination with the government under the umbrella of non-formal education in Lebanon and in collaboration with the government in Niger.

Second, within the HCB remedial education programming, we tested a variety of skill-targeted strategies for improving the HC approach, in an effort to move the needle on children's social-emotional skills and wellbeing. These strategies included SEL activities and games that were implemented in between academic subject-matter transitions during the remedial school day to target specific social and emotional skills. Specifically, we tested the impact of (a) mindfulness activities that targeted stress reactivity and emotion regulation and (b) brain games that targeted specific executive functioning skills such as working memory and behavioral regulation. We also tested embedding a (c) multi-component skill-targeted SEL curriculum – a planned sequence of stand-alone lessons incorporating multiple methods of instructions – within the HC approach. Different SEL activities and curricula as well as combinations of activities were incorporated into the HCB remedial programs in Lebanon and Niger at different times and in different years, as illustrated in Table 2. We collectively refer to these enhanced versions as HCP (Healing Classrooms Plus), while we refer to all versions of remedial programming tested (HCB and HCP) as SEL-infused remedial education programming.

To summarize, in 3EA we provided evidence of the impact of two types of programs in Lebanon and Niger: (a) the impact of access to and dosage of SEL-infused remedial education programs (compared with no access to or reduced dosage of such remedial programming) and (b) the impact of remedial education with access to additional skill-targeted SEL activities, games, and curricula (compared with remedial education programs without access to skill-targeted SEL activities.). These two program foci had a number of implications for our TOCs and research design, as discussed below.

### *Theory(ies) of change*

In order to build a comprehensive understanding of the skills that these innovative programs were expected to improve, and by what mechanisms and with what assumptions, we identified a program-specific TOC for each of the programs evaluated. Specifically, we first identified the most proximal targets that each program was designed to improve. We did so through a collective review and discussion of the key elements of each program with the program developers and implementation partners of the programs (the IRC and Harvard University). As illustrated in Table 3, these immediate, proximal target outcomes were considered *primary outcomes* and key mechanisms of change for each program. For example, based on the findings from the evaluation of HC in the DRC, we identified children's perceptions of their school environments and their literacy and numeracy skills as the primary outcomes of the SEL-infused remedial education program. In contrast, stress response and emotional regulation were the primary outcomes of the mindfulness activities, while executive functioning skills were the primary outcomes of the brain games program.

We then identified more distal target outcomes that were hypothesized to be indirectly impacted by the change of the primary outcomes. We did so by reviewing existing theories and empirical evidence on development across interconnected

domains of children's skills and across contexts. We also explicitly prioritized distal outcomes that have policy implications, such as academic and mental health outcomes. These distal outcomes – along with other outcomes that we did not hypothesize to be impacted by that specific program but were primary or distal outcomes of other programs that were concurrently being evaluated – were included in our TOC and data collection as *secondary outcomes*.

This process resulted in four program-specific TOCs: the HCB remedial education program, two skill-targeted activities (mindfulness, brain games), and the SEL curriculum targeting five distinct SEL skills. The co-construction of these four comprehensive TOCs by the researchers and the practitioners had two main implications for the impact evaluation studies in Lebanon and Niger.

First, these TOCs included a wide range of cognitive, emotional, social, and behavioral skills and experiences. This provided a much more comprehensive picture of intra- and interpersonal developmental processes compared with the reduced-form TOC than we were able to test in the DRC. At the same time, due to the complex research design that evaluated multiple programs concurrently, these partially overlapping – and somewhat disjointed – TOCs included multiple developmental processes about which we did not have a priori hypotheses. This provided a unique opportunity to explore relationships between key developmental processes beyond the hypothesized mechanisms of impacts with underrepresented populations in Lebanon and Niger.

Second, testing these comprehensive TOCs required developing, adapting, and validating context-relevant measures of developmental processes and outcomes that are feasible and appropriate to use in low-resourced settings. As already noted, there is a paucity of measures – particularly social-emotional skills measures – tested for use in non-WEIRD contexts. This is particularly important given how social and emotional skills are defined, prioritized, and manifested varies greatly within and between cultures and contexts (Torrente, Alimchandani, et al., 2015). Therefore, measure development, adaptation, and validation that are fit for the context and purpose – the primary foci of the 3EA measurement work – were critical for meaningful interpretation of the findings from the study (Tubbs Dolan & Caires, 2020).

### *Research design*

The impact evaluations in Lebanon and Niger had two main design features.

The first main design feature was that our two distinct programmatic foci (access to/dosage of remedial education; value added of skill-targeted SEL activities) required a complex experimental design. Testing the impact of access to SEL-infused remedial programming (compared with formal school only) required a business-as-usual control group, while testing the impact of added skill-targeted SEL strategies required at least two groups – those who had access to HCB and those with access to HCP. We achieved this with different designs in Lebanon and Niger given the unique contexts and operational constraints. In Lebanon, in the first year of 3EA programming, we employed a three-arm site-randomized trial, with one arm being a business-as-usual wait-list control group (Tubbs Dolan, Kim, Brown, Gjicali & Aber, 2020). As in the evaluation in the DRC, we leveraged the IRC's lack of financial and human resources to provide services to all beneficiaries at the start of the intervention to create a scientifically and ethically sound research design. However, achieving statistical

**Table 2.** 3EA treatment groups and programs implemented in each group in year 1 and year 2 in Lebanon and Niger.

	Year 1 treatment groups and programs			Year 2 treatment groups and programs	
	Treatment group	Cycle 1	Cycle 2	Treatment group	Full Year
Lebanon	Treatment LB1-A	Remedial support + mindfulness ST-SEL activities	Remedial Support + brain games ST-SEL activities	Treatment LB2-A	Remedial support + ST-SEL curriculum
	Treatment LB1-B	Remedial support	Remedial support	Treatment LB2-B	Remedial support
	Treatment LB1-C	Waitlist (no treatment)	Remedial support + brain games ST-SEL activities		
	Treatment LB1-D	Waitlist (no treatment)	Remedial support		
Niger	Treatment NG1-A	Remedial support + mindfulness ST-SEL activities	Remedial support + brain games ST-SEL activities	Treatment NG2-A	Remedial support + brain Games ST-SEL activities
	Control NG1-A <sup>a</sup>	Waitlist (no treatment)	Waitlist (no treatment)	Control NG2-A <sup>a</sup>	Waitlist (no treatment)
	Treatment NG1-B	Remedial support	Remedial support	Treatment NG2-B	Remedial support
	Control NG1-B <sup>a</sup>	Waitlist (no treatment)	Waitlist (no treatment)	Control NG2-B <sup>a</sup>	Waitlist (no treatment)

Note: ST-SEL = skill-targeted social-emotional learning

<sup>a</sup>Control NG1-A and NG2-A groups consisted of the students in the same school as Treatment NG1A and Treatment NG2-B. Control NG1-B and NG2-B groups consisted of the students in the same school as Treatment NG1-B and Treatment NG2-B

**Table 3.** 3EA Lebanon program primary and secondary outcomes.

Added impact of ST-SEL activities, games, and curricula	HC remedial	Mindfulness ST-SEL activities	Brain games ST-SEL activities	Five-component ST-SEL curriculum
Literacy outcomes	Primary	Secondary	Secondary	Secondary
Numeracy outcomes	Primary	Secondary	Secondary	Secondary
Perceptions of public-school and remedial program environment	Primary	Secondary	Secondary	N/A
Stress response	Secondary	Primary	Secondary	N/A
Emotional and behavioral (dys) regulation	Secondary	Primary	Primary	Primary
Executive functioning	Secondary		Primary	Primary
Emotion knowledge and awareness	N/A	N/A	N/A	Primary
Social cognition and conflict resolution	Secondary	Secondary	Secondary	Primary
Social functioning and competence	N/A	N/A	Secondary	Primary
Mental health	Secondary	Secondary	Secondary	Secondary

Notes: ST-SEL = skill-targeted social-emotional learning. HC = Healing Classrooms. Primary outcomes are outcomes directly targeted by the program key elements and key mechanisms of change for each program. Secondary outcomes are distal outcomes that are hypothesized to be indirectly impacted by the change of the primary outcomes.

power at the site level to field a three-arm trial required the IRC to rapidly scale up their remedial programming in Lebanon, creating implementation challenges and field staff fatigue.

In Niger, the research design was constrained by both the number of schools in which the IRC had funding to operate, as well as the very real possibility that Boko Haram (which literally translates to “western education is forbidden”) could attack schools. In order to attain sufficient power and to ensure the internal validity of our design if treatment schools differentially dropped out, we fielded a design in which pairs of schools were matched on a set of baseline characteristics (Brown, Kim, Tubbs Dolan, & Aber, 2020). We randomized within pairs to implement

either HCB or HCP remedial programming. Then, within schools, over 8,000 students were assessed each year to determine their eligibility for remedial programming. Given that the majority of students met the criteria but there were limited spaces available in remedial classrooms, we created a lottery and randomized eligible individual students to receive remedial programming (or not). Given the size of the student body, however, we were only able to collect the most basic information (grades, attendance, literacy and numeracy skill levels) to test the impact of access to remedial programming.

The second main design feature was that, where possible, we collected longitudinal data about a robust set of developmental

processes and holistic learning outcomes, demographic and ecological risk factors of children and teachers, and (some) program implementation data. We followed the same group of children within years, collecting two to three waves of data in each of the two school years 3EA operated. Because some children remained in the program across years, we also had up to five waves of data for a subset of children.

Taken together, these two design features had several implications.

First, by aligning data collection timing, program implementation cycles, and (in Lebanon) mid-year re-randomization, we were able to experimentally test 10 different research contrasts over 2 years in two countries. These research contrasts enabled us to ask the classic child development and social policy questions about our two programmatic foci. For example, in terms of “what works?” we were able to ask the following questions. (a) Does access to SEL-infused remedial education programming improve children’s school ecologies, academic, social, and emotional processes and outcomes, compared with access to formal school only? (b) Does access to additional skill-targeted SEL activities and curriculum improve children’s targeted social, emotional, and cognitive processes (compared with access to HCB remedial programming only)? Because we were able to examine some of these questions over multiple years and in different countries, we were also able to build confidence in the findings through a form of replication.

Second, by focusing on longitudinal data collection within the experimental trials, we had large samples with which to pose and answer questions about child development in crisis-affected settings from the perspective of developmental psychopathology in context. Such research can inform the design of more contextually appropriate programming and generate evidence to contribute to the global knowledge base for a more representative science of child development. For example, how do experiences and conditions of crisis-affected settings affect the basic processes of *typical* development? To what extent do children’s cultures, communities, and atypical experiences lead to important similarities and differences in their developmental processes and trajectories? In the next part of this paper, we summarize the answers we have arrived at to date and discuss our partnership’s collective interpretation of the results.

### **Key findings to date on the impact of 3EA programming and children’s development**

From both perspectives of child development and social policy and developmental psychopathology in context, we are in the midst of learning more about how to improve children’s academic learning and SEL in conflict and crisis contexts. The results of some of the evaluation studies are still being analyzed and many additional analyses are planned, leveraging large-scale, longitudinal datasets. Nevertheless, the currently available findings build on results from the DRC in ways that are both instructive and intriguing, and we present them here now to guide further research efforts.

#### *What works?*

Overall, the findings from the DRC and early results from Lebanon and Niger provide mounting evidence that the HC approach – whether embedded in formal school curricula (as in the DRC) or as remedial education programming (as in

Lebanon and Niger) – can work to improve children’s perceptions of their schools and their basic academic skills.

Firstly, recall that, in the DRC, access to HC improved children’s perceptions of the safety and supportiveness of their schools after 1 year of programming, but not after 2 years. We found similar results in Lebanon. A half-year of access to the SEL-infused remedial programming – including both HCB and HCP – increased children’s perceptions of the safety and supportiveness of their Lebanese public school environment. A full-year of access to the same remedial programming, however, did not further improve children’s perceptions (Tubbs Dolan *et al.*, 2020).

Secondly, recall from the evaluation in the DRC that access to HC programming significantly improved children’s literacy and numeracy skills, but that the impacts were not large enough to put children on the path to literacy and numeracy by the end of primary school. In the first year of 3EA in Niger, access to a full year of SEL-infused remedial education programming – including both HCB and HCP – improved both Nigerian refugee and Nigerien local students’ literacy and numeracy scores (effect size = .22 to .28). These positive impacts were consistent across children of different gender, grade level, refugee status, and baseline literacy and numeracy scores (Brown *et al.*, 2020). In Lebanon, we were able to evaluate the impact of SEL-infused remedial programming on a more comprehensive set of academic skills. We found that short-term (16 weeks) access to SEL-infused remedial education programming improved some of the discrete basic literacy and numeracy skills (e.g., letter recognition and number identification), but not higher order literacy (e.g., oral passage reading, reading comprehension) and numeracy (e.g., word problems) skills, or global literacy and numeracy competency (Tubbs Dolan *et al.*, 2020).

As in the DRC, we suspect that implementation and dosage played a role in the consistency and size of the impacts on perceptions of school environment and academic outcomes. Unlike in the DRC study, we have some data that will allow us to unpack these findings. It is illustrative, for example, that the average attendance rate in remedial programming was 50% in Lebanon and 64% in Niger. These low rates of attendance represent the numerous challenges of and competing priorities for engaging in education in crisis-affected settings (Brown *et al.*, 2019; Keim & Kim, 2019) and they highlight that context-relevant implementation strategies must be developed to ensure impacts are consistent, sustainable, and meaningful among populations where frequent attendance is not possible.

#### *What doesn’t work (yet)?*

Overall, we found that moving the needle on children’s social and emotional outcomes in crisis contexts remains elusive. In Lebanon, we did not find that any of the three types of SEL programs tested – mindfulness, brain games, and five-component SEL – had conclusive, significant impacts on children’s social and emotional skills. These disappointing results may be partially due to the relatively small dosage and treatment contrasts, as these skill-targeted SEL activities and curricula were designed to be short and quick activities implemented within the larger HCB remedial education programs. Combined with the above-described attendance and quality of implementation issues that plague programs in humanitarian settings, these programs – implemented for a relatively short duration – were not sufficient to make meaningful improvements in children’s social and emotional skills. We did observe that some of the skill-targeted SEL programs had potentially promising positive impacts on primary

social and emotional outcomes, but the impacts were not large enough or consistent enough to provide statistically conclusive evidence. In Niger, there was also evidence that access to additional mindfulness and brain games activities improved students' grades in their public schools (Brown et al., 2020). We will continue to explore potential heterogeneity of impacts to identify how, for whom and under what conditions these skill-targeted programs may improve children's academic and social and emotional outcomes.

### *Child development in crisis-affected settings*

Leveraging the data on the wide range of cognitive, social, emotional, and behavioral processes that are hypothesized to be the primary and secondary outcomes of the 3EA programs, we were also able to examine what child development looks like and what are the predictors of adaptive and maladaptive development in Lebanon and Niger. This gave us an opportunity to explicitly test some of the hypothesized relations between children's developmental processes and outcomes that we hypothesized in the TOCs in the contexts for which the programs were designed.

We have only begun to unravel such complexity and nuances. In the first developmental study to result from the 3EA data (Kim, Brown et al., 2020), we found that Syrian refugee children face many pre-, peri-, and post-migration risks, some of which are especially salient among refugee populations (e.g., being assigned to a lower grade than is normative for their age) and some of which are nearly universal risk factors (e.g., poor health). These risks are associated with decrements in children's ability to regulate themselves cognitively and behaviorally, and to achieve literacy and numeracy skills (Kim, Brown et al., 2020). We also found that children's regulatory skills, including cognitive (executive function) skills and behavioral regulation skills, are key predictors of children's academic learning. These findings replicate research from WEIRD contexts that emphasize the importance of SEL skills for children's ability to learn. These regulatory skills were also primary targets of some of the 3EA skill-targeted SEL programs and, as such, these findings provide support for the proposed TOCs of these programs. Further analysis on how these developmental processes and risks dynamically interact over time and across different cultural and sociopolitical contexts will help us to better understand the experiences of children growing up in crisis-affected settings and to identify more effective ways to support children's positive development.

### *Lessons learned towards a science for action: 3EA*

As with the 3EA research findings, we are currently in the midst of processing what we have learned to date from 3EA about how to build a science for action for children in crisis contexts. We offer our mid-course reflections on these learnings in the context of the four themes introduced at the start of this paper.

#### *Partnerships*

Our experiences with the 3EA initiative reinforced our finding that research-practice partnerships can be an effective and ethical strategy for generating and using evidence in crisis contexts. Our experiences with the IRC in 3EA further highlighted the power of deep respect in partnerships to motivate the mutual capacity building required for turning science into action. Colleagues at the IRC now regularly ask questions about effect sizes, treatment contrasts, and heterogeneity of impact, while we at NYU-TIES obsess over what the findings mean for practice and how to

communicate findings to practitioners and policymakers clearly and intuitively. These are critical skills for researchers and practitioners to learn if we truly want to move towards evidence-informed programming for children in crisis contexts.

However, building these skills – as individuals and as organizations – has a price. Academic institutions tend to place a higher premium on publications in peer-reviewed journals than on communication to non-technical audiences or the soft skills required for sustained partnership. Many humanitarian organizations believe that funding and time is better spent on programming than on research. Enabling the individual and organizational pursuit of such partnerships thus requires seismic shifts in sectoral norms and imperatives. Even if pursued, maintaining healthy partnerships requires consistency of staff and resources to ensure coordination and sustained capacity. Fortunately, we were able to form strong relationships with key donors that provided the multi-year funding required for such sustained staffing and capacity. By moving from traditional output/accountability models to joint learning/continuous improvement models, donors can encourage a culture of sharing what works and what hasn't worked yet (as reflected in this paper). This information is critical for making decisions about how to invest scarce program and research resources in crisis contexts.

We also recognize that we must do more to ensure that the voices and engagement of local communities, researchers, NGOs, and governments are front and center in research and in partnerships. We were able to begin to build such relationships through our 3EA measurement research (Tubbs Dolan & Caires, 2020). However, as we reflect in the wake of the killing of George Floyd and the protests for racial and social justice, we recognize that we have not done enough to ensure that our interactions, assumptions, methods, and partnerships actively contribute to anti-racism and decolonization in the contexts in which we work. We commit to doing better.

#### *Conceptual frameworks*

In the evaluation in the DRC, we lived with the limitations of a reduced-form TOC. In 3EA, we have the opposite issue: we are living with both the benefits and challenges of multiple (at least four) comprehensive TOCs. We say "benefits" because they allowed us to specify and measure key constructs necessary for answering questions within the perspectives of child development and social policy and developmental psychopathology in context. These include measures of the implementation of the skill-targeted SEL activities and curricula, developmental processes, ecological risk factors, and holistic learning outcomes. We say "challenges" because it is difficult to take stock of our findings across all of the constructs in all of the different TOCs. While we are perhaps better poised to answer *why* we have null results, we are limited in our ability to determine *what to do about it* if we look contrast by contrast, construct by construct.

We now need to move towards aligning our findings – and those being generated by other stakeholders in both the education and child protection sectors – within a conceptual framework. Conceptual frameworks do not focus on the specific content of strategies and approaches; rather, they specify the expected relationships among broad components of education systems and programs and children's developmental skills and outcomes. In doing so, they provide a heuristic against which diverse stakeholders can map and organize specific findings, comprehensively examine and process the implications of those findings, and communicate findings with others (Blyth, Jones, & Borowski, 2018).

Such organization and communication are necessary for ensuring that the limited resources available for research in crisis contexts result in a coherent evidence base from which it is possible to draw recommendations for program and policy decision making (Tubbs Dolan, 2018).

#### *Methods and measures*

While we made significant improvements over the methods and measures used in the evaluation in the DRC, we learned in 3EA that we still had a long way to go.

First, while we worked with the IRC to create or modify existing administrative data systems to track children over time, the process for doing so was error prone and intensive in terms of time and resources for the scale of data collection to which we aspired: it took thousands of hours to clean and organize 1 year of data from one of the 3EA countries so that we could confidently use it for analyses. Some of the issues we encountered can be solved before and during collection itself: for example, through better enumerator training and monitoring of enumerator quality (Brown & Ngoga, 2019). However, more broadly, what is required is investments in the systems and capacity for data cleaning, managing, and curating (Anker & Falek, 2020). In addition, even once cleaned, methodological advances are needed to improve missing data imputation and causal inference in contexts with complex migration patterns and selection processes (Montes de Oca Salinas, 2020).

Second, we made progress in measuring CHILD outcomes and program implementation, including through performance-based (Ford, Kim, Brown, Aber, & Sheridan, 2019), scenario-based (Kim & Tubbs Dolan, 2019), teacher-report (Kim, Gjicali, Wu, & Tobbs Dolan, 2020), and assessor-report measures. Measures were either adapted or assembled from WEIRD settings (when necessary due to time constraints) or specifically developed with the population and setting in mind (when additional resources were available). All measures, however, required extensive analytical testing that resulted in either (a) adjustments to meet the psychometric standards necessary for impact evaluation or (b), in some rare instances, significant adaptation or being deemed unreliable for the study. These experiences reinforced to us that strong measures are not instantly created: they are developed and refined over time as evidence accumulates with new trials, in different contexts, or for distinct purposes. Identifying stronger measures is like trying to identify what programs work best for children. It is hard to draw broad conclusions from one or two evaluations of programs in different contexts – dozens of trials are needed to have confidence that the program is really working and achieving what is intended. The same is true for measures. Having a ready supply of measures that can be used to generate evidence in crisis contexts requires a community of users who are empowered and committed to using the measures over time and reporting back on psychometric properties (Tubbs Dolan & Caires, 2020).

#### *Joint reflection*

Our initial goal was to rapidly prototype effective SEL programs. We aimed to do so by consecutively testing three different types of SEL programs over a 2-year period, making decisions about the next year's programs based on usable and reliable evidence from the first year's evaluation study. However, this approach, combined with large-scale data collection and implementation, turned out to be challenging and stressful for both organizations. While we were able to produce preliminary evidence to inform some of the programmatic and research decisions in this time

frame, some of the preliminary findings we have shared for decision making turned out to be inconsistent with the final results, and often insufficient to be used as bases for program decisions. We found that if the practice partner needs rapid feedback, a large-scale randomized controlled trial is not the right answer. If we could do it again, we would conduct different types of studies for different purposes over a longer time frame. First, we would conduct a series of targeted, rapid prototyping, and feasibility studies for program development and measures validation during the inception period. Second, we would be sure to build in time to inform revision of program design. Third, once we had built confidence in a program through short-term, small-scale, and iterative feasibility studies, adequately powered randomized controlled trials would be needed to generate evidence of effectiveness of the program at scale. For this type of iterative research–practice feedback loop to work, a longer term commitment of and partnerships with funders would also be critical. We thus continue.

#### **Building a Developmental Science for Action**

In this final section of the paper, we draw on our experiences in research–practice–policy partnerships over the last decade to identify the essential features of a global developmental science for action that strives to understand and enhance children's learning and development in vulnerable families, communities, and countries. What do we mean by a developmental science for action? We mean nothing more, but nothing less, than a science that can inform and help guide the actions of key forces in children's lives – parents and teachers at the microsystem level, program providers and service institutions at the mesosystem level, and policymakers at the exosystem level of human ecology. Said another way, a science for action produces knowledge that contributes to decisions on behalf of children and their families that are better than if the knowledge was not available. For our purposes in this paper, we focused on science to inform and guide program providers (like the IRC and local NGOs) and policymakers (particularly, governments and donors) about what works to improve the learning and development of children in conflict and crisis contexts. The essential features we emphasize here in these concluding remarks are neither definitive nor complete. Others will find important problems with our formulations and important errors of omission and commission in this set of features. However, we do hope to provoke thought among scholars of the child development and social policy and developmental psychopathology in context perspectives about what is required to build a global developmental science for action.

#### *Attentive to the basics of an experimental social-behavioral-developmental science*

First and obviously, a global developmental science for action must be attentive to the basics of all the experimental prevention and intervention sciences in order to answer a question of paramount importance to program providers and policymakers: what works? This is difficult under any circumstance, but especially when conducting what Donald Campbell called field experiments to enhance children's learning and development in countries of conflict and crisis. As Campbell wrote "Measuring the effects of complex politically designed ameliorative programs involves all the problems of inference found in measuring the effects of a conceptually pure treatment variable – all and more" (Campbell, 1998, p. 45). As we emphasized throughout

this paper, issues of causal inference (what works?) are paramount, but so are issues of mechanisms of change (how?), heterogeneity of impact (for whom and under what conditions?), and generalizability (will it work again at other times, with new populations, and in new places?). Through our research, we wanted to learn if HC worked not only in formal schools in the DRC but also in after-school programs in Niger or in remedial support programs in Lebanon. Could we generalize the findings about the positive impacts of 1 year of HC in the DRC to the potential impacts of 1 year of HC in other conflict- and crisis-affected countries? Strictly speaking, of course not; practically speaking, we hope so! Are we on firmer ground generalizing converging findings not only from DRC but now also from Niger and Lebanon beyond these countries to inform what might work in other conflicts, crises, countries? Again, strictly speaking, no; and yet we must. Action in the humanitarian and international development sectors requires it. As the Nobel Laureate physicist Feynman (1964, 41:09) observed “The moment that you make a proposition about a region of experience that you haven’t directly seen, you must be uncertain. But we must always make statements about the regions we haven’t seen, or it’s no use in the whole business.” A global developmental science for action hopes to generalize beyond limited studies to make statements about what may work for regions of experience we haven’t directly seen.

### *Cumulative, revisable, and usable science*

Second, a developmental science for action must be cumulative, revisable, and usable. To state the obvious, no single impact evaluation of an intervention provides an adequate basis for generalizable knowledge about what works. However, research–practice–policy partnerships can mount theory-driven and culturally valid evaluations that in turn can lead to replication studies, systematic reviews, meta-analyses of multiple studies, and theory-based research syntheses. Via such processes, knowledge about what works in a particular domain can cumulate. Such cumulative knowledge is, we believe, a stronger basis to use such knowledge for action on behalf of children. This is the history of research on what works to improve children’s academic learning and SEL in the USA (Durlak et al., 2011). After more than 20 years and over 200 studies involving over 200,000 children, we have (relatively more) generalizable knowledge of what works to improve children’s academic outcomes and social and emotional wellbeing in the USA. This cumulative knowledge has led to changes in SEL practices, programs, and systems in the USA (Dusenbury et al., 2015; Fagan, Hawkins, & Shapiro, 2015; Zaslow, Mackintosh, Mancoll, Mandell, & Mart, 2015). The history of building a cumulative, revisable, and usable knowledge base in the USA brings us back to an earlier challenge we first faced in the DRC. The resources and time to conduct many high-quality impact evaluations about what works to enhance the learning and development of children in conflict- and crisis-affected countries are not currently available – they must become available to achieve generalizable knowledge of what works. The political economy of generating knowledge about what works in low-income, conflict- and crisis-affected countries must change to build a global developmental science for action. How?

### *Grounded in ethical principles and normative goals*

The only way forward to an adequate developmental science for action is to invest more and invest smartly in research on children

in conflict and crisis contexts – on what basis can we do so? We argue that the world has been developing such an ethical and normative basis for centuries. As noted earlier, these efforts entered a new stage with the passage of the CRC in 1989 (UNCRC, 1989) and have accelerated further in the following decades with the formulation of the Millennium Development Goals (covering the period from 1990 to 2015) and their successor, the Sustainable Development Goals (SDGs, covering the period from 2015 to 2030) (UNDESA, 2015). Nearly every country in the world has signed on to both the CRC and the SDGs. In principle, all the nations of the world now formally agree to a set of rights for children (especially articles 28 and 29 of the CRC regarding the right to an education) and a set of development goals (SDG4, which, among other things, pledges that by 2030 all children will receive free quality primary and secondary education) to guide and evaluate our actions on behalf of children. While stating rights and pledging to goals is no guarantee that such rights and goals will be achieved, they make it possible for developmental scientists to join others in advocating for full implementation of the CRC and the SDGs. In other words, changes in the political economy of a global developmental science for action cannot rest solely with activity within the scientific community. A global developmental science for action must be grounded in legal rights and normative goals, which are the practical ends to which the means of a developmental science for action must serve. Such an engaged science will not be welcomed by all developmental scientists. Some will want science to remain “pure,” uncontaminated by value preferences of the scientists, but what if all the world’s governments have formally endorsed such rights and norms? These seem to us a firm foundation to pursue both the child development and social policy and the developmental psychopathology in context perspectives championed over decades by Edward Zigler.

**Acknowledgments.** The authors wish to thank our practice partner, the International Rescue Committee (IRC), for its dual commitments to high-quality humanitarian aid and to generating and using rigorous evidence on what works in promoting children’s learning and development in conflict- and crisis-affected countries. We especially wish to thank Dr Jeannie Annan (head of the research, evaluation, and learning unit of the IRC), Dr Sarah Smith (head of the education unit of the IRC), the IRC country directors and country heads of education in the DRC, Niger, and Lebanon, and the numerous other IRC colleagues at central office, regional, and country levels who made this research possible. We especially want to thank the hundreds of school directors and teachers and the thousands of children in the Democratic Republic of Congo, Niger, and Lebanon who participated in this research over the last decade. Finally, we thank our many co-authors on the studies reviewed in this paper and our exceptional colleagues at New York University, especially the staff of Global TIES for Children and the Institute for Human Development and Social Change.

**Financial Statement.** The impact evaluation of HC in the DRC was funded by the United States Agency for International Development under cooperative agreement 623-A-10-00023 and by the NoVo Foundation. The program implementation and impact evaluations of HC in Niger and Lebanon were funded by Dubai Cares, Spencer Foundation, PRM, the UK Economic and Social Research Council (ESRC), the UK Department for International Development (DFID), and two anonymous donors. We also thank the New York University Abu Dhabi Research Institute for core funding to the New York University Global TIES for Children center, which provides support for staff and activities that directly contributed to the research presented here. In addition, we thank Dubai Cares and ESRC-DFID for supporting continued analyses of the datasets described in this manuscript.

**Conflicts of Interest.** The authors of this paper report no formal conflicts of interest. We do acknowledge that the joint research–practice partnership model we used to conduct this research involves very close collaboration in the co-construction of the research questions for and interpretation of data from these impact evaluations. All final decisions about the design and implementation of the interventions were made by the IRC. All final decisions about the research questions, design, analysis, and interpretation of the data were made by the authors.

## References

- Aber, J. L., Bishop-Josef, S. J., Jones, S. M., McLearn, K. T., & Phillips, D. A. (Eds.). (2007). *Child development and social policy: Knowledge for action* (pp. xvii, 311). Washington, DC: American Psychological Association.
- Aber, J. L., Torrente, C., Starkey, L., Johnston, B., Seidman, E., Halpin, P., ... Wolf, S. (2017). Impacts after one year of “Healing Classroom” on children’s reading and math skills in DRC: Results from a cluster randomized trial. *Journal of Research on Educational Effectiveness*, 10, 507–529. doi:10.1080/19345747.2016.1236160
- Aber, J. L., Tubbs, C., Torrente, C., Halpin, P. F., Johnston, B., Starkey, L., ... Wolf, S. (2017). Promoting children’s learning and development in conflict-affected countries: Testing change process in the Democratic Republic of the Congo. *Development and Psychopathology*, 29, 53–67. doi:10.1017/S0954579416001139
- Anker, P., & Falek, I. (2020). *Data pipeline: From chaos to catalyst* [internal presentation]. New York, NY: Global TIES for Children.
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63, 602–614. doi:10.1037/0003-066X.63.7.602
- Betancourt, T. S., & Williams, T. (2008). Building an evidence base on mental health interventions for children affected by armed conflict. *Intervention*, 6, 39–56. doi:10.1097/WTF.0b013e3282f761ff
- Blyth, D. A., Jones, S., & Borowski, T. (2018). *SEL frameworks – What are they and why are they important*. Retrieved from <https://measuringSEL.caselo.org/wp-content/uploads/2018/09/Frameworks-A.1.pdf>
- Brown, L., Kim, H. Y., Annan, J., & Aber, J. L. (2019). *SEL functioning in crisis-contexts: The role of gender, grade level, and school attendance in Sierra Leone*. Washington, DC: Society for Research in Educational Effectiveness.
- Brown, L., Kim, H. Y., Tubbs Dolan, C., & Aber, J. L. (2020). *Remedial programming and skill-targeted SEL in low-income contexts: Experimental evidence from Niger*. Unpublished manuscript.
- Brown, L., & Ngoga, E. (2019). Examining the enumerator effect: Improving data quality through enumerator observation. *Network for International Policies and Cooperation in Education and Training Special Issue*, 2, 59–62.
- Burde, D., & Linden, L. L. (2013). Bringing education to Afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, 5, 27–40. doi:10.1257/app.5.3.27
- Campbell, D. T. (1998). The experimenting society. In W. N. Dunn (Ed.), *The experimenting society: Essays in honor of Donald T. Campbell* (Policy Studies Review Annual, Vol. 11, pp. 35–68). New Brunswick, NJ: Transaction Books.
- Cicchetti, D., & Aber, J. L. (1998). Contextualism and developmental psychopathology. *Development and Psychopathology*, 10, 137–141.
- Cummings, E. M., Goeke-Morey, M. C., Merrilees, C. E., Taylor, L. K., & Shirlow, P. (2014). A social-ecological, process-oriented perspective on political violence and child development. *Child Development Perspectives*, 8, 82–89. doi:10.1111/cdep.12067
- DeStefano, J., Moore, A.-M. S., Balwanz, D., & Hartwell, A. (2007). *Reaching the underserved: Complementary models of effective schooling*. Retrieved from <https://eric.ed.gov/?id=ED505689>
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432. doi:10.1111/j.1467-8624.2010.01564.x
- Dusenbury, L. A., Newman, J. Z., Weissberg, R. P., Goren, P., Domitrovich, C. E., & Mart, A. K. (2015). The case for preschool through high school state learning standards for SEL. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 532–548). New York, NY: Guilford Press.
- Fagan, A. A., Hawkins, J. D., & Shapiro, V. B. (2015). Taking SEL to scale in schools: The role of community coalitions. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 468–481). New York, NY: Guilford Press.
- Feynman, R. P. (1964). *Messenger Feynman great conservation principles*. Retrieved from <https://www.youtube.com/watch?v=cxSF09xrAIE>
- Ford, C. B., Kim, H. Y., Brown, L., Aber, J. L., & Sheridan, M. A. (2019). A cognitive assessment tool designed for data collection in the field in low- and middle-income countries. *Research in Comparative and International Education*, 14, 141–157. doi:10.1177/1745499919829217
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Gove, A., & Wetterberg, A. (Eds.) (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press.
- Halpin, P. F., & Torrente, C. (2014). *Measuring critical education processes and outcomes: Illustration from a cluster randomized trial in the Democratic Republic of the Congo*. Society for Research on Educational Effectiveness. Retrieved from <https://eric.ed.gov/?id=ED562783>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–83. doi:10.1017/S0140525X0999152X
- Jordans, M. J. D., Tol, W. A., Komproe, I. H., & Jong, J. V. T. M. D. (2009). Systematic review of evidence and treatment approaches: Psychosocial and mental health care for children in war. *Child and Adolescent Mental Health*, 14, 2–14. doi:10.1111/j.1475-3588.2008.00515.x
- Keim, A., & Kim, H. Y. (2019). *Student- and classroom-level predictors of refugee student attendance trajectories in Lebanon: Implications for interventions*. Baltimore, MD: Society for Research in Child Development.
- Kim, H. Y., Brown, L., Tubbs Dolan, C., Sheridan, M., & Aber, J. L. (2020). Post-migration risks, developmental processes, and learning among Syrian refugee children in Lebanon. *Journal of Applied Developmental Psychology*, 69, 101142. doi:10.1016/j.appdev.2020.101142
- Kim, H. Y., Gjicali, K., Wu, Z., & Tobbs Dolan, C. (2020). *Teachers’ Observations of Learners’ Social and Emotional Learning (TOOLSEL): Psychometric evidence for use in program evaluation in education in emergencies*. Unpublished manuscript.
- Kim, H. Y., & Tubbs Dolan, C. (2019). *Social Emotional Response and Information Scenarios (SERAIS): Evidence on construct validity, measurement invariance, and reliability in use with Syrian refugee children in Lebanon*. Unpublished Manuscript. doi:10.13140/RG.2.2.23945.60007
- Luthar, S. S., Burack, J. A., Cicchetti, D., & Weisz, J. R. (Eds.) (1997). *Developmental psychopathology: Perspectives on adjustment, risk, and disorder* (pp. xxi, 618). Cambridge: Cambridge University Press.
- Masten, A. S., & Narayan, A. J. (2012). Child development in the context of disaster, war, and terrorism: Pathways of risk and resilience. *Annual Review of Psychology*, 63, 227–257. doi:10.1146/annurev-psych-120710-100356
- McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, doi:10.3102/0034654314553127
- Montes de Oca Salinas, M. (2020). *Practical imputation strategies to address missing data from complex social studies*. Unpublished manuscript.
- Olenik, C., Fawcett, C., & Boyson, J. (2013). *State of the field report: Examining the evidence in youth workforce development*. Washington, DC: USAID.
- Prunier, G. (2009). *From genocide to continental war: The “Congolese” conflict and the crisis of contemporary Africa*. London: Hurst & Co.
- Torrente, C., Aber, J. L., Starkey, L., Johnston, B., Shivshanker, A., Weisenhorn, N., ... Dolan, C. T. (2019). Improving primary education in the Democratic Republic of the Congo: End-line results of a cluster-randomized wait-list controlled trial of Learning in a Healing Classroom. *Journal of Research on Educational Effectiveness*, 12, 413–447. doi:10.1080/19345747.2018.1561963
- Torrente, C., Alimchandani, A., & Aber, J. L. (2015). International perspectives on social and emotional learning. In J. A. Durlak, C. E. Domitrovich, R.

- P. Weissberg & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 566–588). New York, NY: Guilford Press.
- Torrente, C., Johnston, B., Starkey, L., Seidman, E., Shivshanker, A., Weisenhorn, N., ... Aber, J. L. (2015). Improving the quality of school interactions and student well-being: Impacts on one year of a school-based program in the Democratic Republic of Congo. *Journal of Education in Emergencies, 1*, 48–91.
- Tubbs Dolan, C. (2018). *Methods for rigorously and responsively improving children's holistic learning and development in crisis contexts: Towards an evidence-based education in emergency field*. Doctoral dissertation, New York University.
- Tubbs Dolan, C. T., & Cairns, R. (2020). The assessment of vulnerable children's social-emotional skills in MENAT. In *Social and emotional learning in the Mediterranean* (pp. 159–176). Leiden, The Netherlands: Brill Sense.
- Tubbs Dolan, C., Kim, H. Y., Brown, L., Gjicali, K., & Aber, J. L. (2020). *Supporting Syrian refugee children in formal education systems: A cluster randomized controlled trial of non-formal remedial and mindfulness programs in Lebanon*. Unpublished manuscript.
- Tubbs Dolan, C., & Weiss-Yagoda, J. (2017). *An EiE research-practice partnership: Learning to improve academic and social-emotional outcomes*. Retrieved from [https://static1.squarespace.com/static/583af1fb414fb5b3977b6f89/t/59bdb9f76f4ca3d8bb13c7ed/1505606138131/13\\_PromisingPractices\\_\\_IRC\\_WEB.pdf](https://static1.squarespace.com/static/583af1fb414fb5b3977b6f89/t/59bdb9f76f4ca3d8bb13c7ed/1505606138131/13_PromisingPractices__IRC_WEB.pdf)
- UNCRC (United Nations Convention on the Rights of the Child). (1989) <https://www.unicef.org/child-rights-convention#Learn>
- UNDESA (United Nations Department of Economic and Social Affairs). (2015). *Sustainable Development Goals (SDGs)*. Retrieved from <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- UNESCO (United Nations Educational, Scientific and Cultural Organization). (2013). Half of all out of school children live in conflict-affected countries. Retrieved from <https://en.unesco.org/news/unesco-half-all-out-school-children-live-conflict-affected-countries>
- UNESCO (United Nations Educational, Scientific and Cultural Organization). (2017). Global education monitoring report 2017/18: Accountability in education: meeting our commitments. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000259593>
- UNICEF (United Nations Children's Fund). (2016). Nearly a quarter of the world's children live in conflict or disaster-stricken countries. UNICEF Press Release, 9 December 2016. Retrieved from <https://www.unicef.org/press-releases/nearly-quarter-worlds-children-live-conflict-or-disaster-stricken-countries>
- Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., Bokkel, I. t., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis. *Cambridge Journal of Education, 46*, 347–376. doi:10.1080/0305764X.2016.1195791
- Winthrop, R., & Kirk, J. (2008). Learning for a Bright Future: Schooling, Armed Conflict, and Children's Well-Being. *Comparative Education Review, 52*, 639–661. doi:10.1086/591301
- Wolf, S., Torrente, C., Frisoli, P., Weisenhorn, N., Shivshanker, A., Annan, J., & Aber, J. L. (2015). Preliminary impacts of the "Learning to Read in a Healing Classroom" intervention on teacher well-being in the Democratic Republic of the Congo. *Teaching and Teacher Education, 52*, 24–36. doi:10.1016/j.tate.2015.08.002
- World Bank. (2018). *World development report 2018: Learning to realize education's promise*. Washington, DC: World Bank.
- Wu, Z., & Brown, W. (2020). *Understanding implementation of SEL kernel activities and school attendance*. Unpublished manuscript.
- Wuermli, A. J., Tubbs, C. C., Petersen, A. C., & Aber, J. L. (2015). Children and youth in low- and middle-income countries: Toward an integrated developmental and intervention science. *Child Development Perspectives, 9*, 61–66. doi:10.1111/cdep.12108
- Zaslow, M., Mackintosh, B., Mancoll, S., Mandell, S., & Mart, A. K. (2015). Federal policy initiatives and children's SEL. In C. E. Domitrovich, R. P. Weissberg & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 549–565). New York, NY: Guilford Press.
- Zigler, E., & Glick, M. (2001). The developmental approach to adult psychopathology. *The Clinical Psychologist, 54*, 2–11.
- Zigler, E., & Hall, N. W. (2000). *Child development and social policy: Theory and applications* (pp. xxii, 338). New York, NY: McGraw-Hill.
- Zigler, E., & Styfco, S. J. (2010). The hidden history of Head Start. In *The hidden history of Head Start*. Oxford: Oxford University Press. <https://www.doi.org/10.1093/acprof:oso/9780195393767.001.0001>