

# Evaluating Program Enhancement Strategies for Remedial Tutoring: A Cluster-Randomized Control Trial With Syrian Refugee Students in Lebanon

Lindsay Brown 

New York University Global TIES for Children

Kalina Gjicali

Meta, New York University

Ha Yeon Kim 

Carly Tubbs Dolan 

New York University Global TIES for Children

Paul Frisoli

The LEGO Foundation

Mahmoud Bwary

United Nations Children's Fund (UNICEF)

J. Lawrence Aber

New York University Global TIES for Children

*Despite widespread enthusiasm for remedial education programming with refugee populations, there is little rigorous evidence on how to design and implement such programs. We employ a cluster-randomized design of non-equivalent treatment groups to test the impact of access to two types of program enhancement: longer program duration and the addition of skill-targeted social and emotional learning (SEL) activities for Syrian refugees enrolled in Lebanese public schools. We find that, compared to 10 weeks of programming, 26 weeks marginally increases students' literacy skills ( $ES = 0.04$ ) and significantly improves behavioral regulation ( $ES = 0.31$ ), but students reported less positive perceptions of their public school environment ( $ES = -0.83$  to  $-0.89$ ) and remedial tutoring site ( $ES = -0.15$  to  $-0.24$ ). We also find that the addition of skill-targeted SEL activities to 26 weeks of programming results in higher student reports of school-related stress compared to programming without skill-targeted activities ( $ES = 0.21$ ). Implications for program and policy are discussed.*

**Keywords:** *child development, cognitive processes/development, experimental research, international education/studies, program evaluation, randomized controlled trial, refugee education, remedial education, social-emotional learning*

As of 2015, an estimated 617 million school-aged children—a slight majority of school-aged children worldwide—lacked minimum proficiency in reading and mathematics (UNESCO, 2017). Students displaced by conflict and crisis—arguably among the most vulnerable—fare even worse. More than half of 7.1 million refugee students around the world are out of school completely, temporarily, or

permanently. Those who *are* able to attend schools lag behind peers in basic literacy and numeracy skills, and enrollment rates decrease precipitously over time (OECD, 2017). Beyond academics, pre- and postmigration experiences can leave refugees with lingering trauma that can result in cognitive and mental health challenges (Kim et al., 2020; Mendenhall et al., 2017).



To stem the learning loss experienced by children in crises such as armed conflicts and pandemics, and to buoy their academic proficiency, the majority of recently surveyed ministries of education report intentions to introduce and/or expand remedial education programming (UNESCO et al., 2020). Remedial education programming is generally defined as a short-term increase in instructional time or targeted academic support to students whose proficiencies in content or skill are below expected levels for grade. But widespread enthusiasm for remedial programming's implementation precedes strong evidence for its effectiveness in non-Western contexts. A small but growing evidence base from middle-income contexts demonstrates that the provision of remedial educational programming can improve learning (Banerjee et al., 2007; Saavedra et al., 2019), but there is almost no evidence from vulnerable populations in crisis contexts where students are arguably most in need of such services. Moreover, there is little actionable evidence on how to design such programs, and few studies investigate social and emotional outcomes alongside academic ones.

There are several reasons to be cautious in simply extending the promising yet nascent evidence base on remedial education to the refugee child population in low- and middle-income countries. First, most of the current evidence on remedial programming investigates programs implemented in formal schooling contexts. Such examples include the “Balsakhi” or “friend of the child” program in India where academically struggling students in municipal schools were tutored during school hours by young women from the local community (Banerjee et al., 2007); an after-school science tutoring program in municipal schools in Lima, Peru, targeting low-achieving third graders (Saavedra et al., 2019); and the support of a “Mobile Pedagogical Tutor”—recent college graduates with short-term training in math and literacy tutoring—in Mexico (Agostinelli et al., 2019). Yet governments in regions affected by conflict and crisis often shoulder a disproportionately large burden of schooling vulnerable populations of refugees and displaced persons (Ruauadel & Morrison-Métois, 2017). In many cases, governments are struggling to provide even basic education services to an influx of displaced students, let alone provide quality supplementary educational programs. In Lebanon, for example, an influx of refugees fleeing conflict in neighboring Syria led to the doubling of the student population within five years (Ministry of Education and Higher Education, 2018). This rapid expansion of students created immense challenges for delivering quality education services. Under such circumstances, service provision by non-state actors like NGOs presents a complementary and potentially valuable method of providing additional educational support.

Second, *providing* access to high-quality services is only half of the learning equation. Contexts of crisis and conflict where refugee students reside are often marked by a highly mobile population, household economic stressors, and

unpredictable security threats (Human Rights Watch, 2016). Such factors can limit student attendance, ultimately decreasing the dosage of programming that students receive (Brown et al., 2023; Brudevold-Newman et al., 2023). Unstable contexts may, therefore, require unique implementation considerations—such as a longer duration of programming—in order to achieve impacts comparable to stable contexts.

Last, early life experiences of vulnerable student populations—such as extreme poverty, violence, or displacement—can profoundly impact children's cognitive, social, and emotional development (Kim et al., 2020; Reed et al., 2012). Academically, students with such adverse early life experiences may require different or additional support in order to accomplish skills that precede academic success, such as successfully deploying attention or processing new information. Socially, students who have experienced trauma are more likely to struggle to regulate their emotions and/or to establish positive relationships with peers and/or adults (Keresteš, 2006; Masten & Narayan, 2012). These differences likely require an educational approach that prioritizes learning within a safe, predictable, and supportive classroom environment and may necessitate activities that support building specific social and emotional skills (e.g., executive function, behavioral regulation skills).

The present study examines two approaches to strengthen the design and impact of remedial tutoring programming for refugee children using a cluster-randomized design of non-equivalent treatment groups with Syrian refugee children enrolled in Lebanese public schools during school year 2016–2017. An earlier, short-term evaluation of the program demonstrated positive impacts on refugee children's school adaptation outcomes (Tubbs Dolan et al., 2022). In this study, we use non-equivalent treatment groups to investigate two important components of remedial programming: duration and skill-targeted SEL activities. First, we ask whether longer duration of a remedial tutoring program infused with social and emotional learning (SEL) principles (10 versus 26 weeks) impacts Syrian refugee students' academic and social and emotional outcomes. Second, we investigate whether the addition of skill-targeted SEL activities to the remedial tutoring program over the full 26 weeks impacts students' academic and social and emotional outcomes compared to the base tutoring program alone.

### *Duration of Remedial Programming*

Despite the potential of remedial programming, high-quality implementation remains elusive, particularly in fragile contexts that host the majority of the world's refugees (Brudevold-Newman et al., 2023). For example, contextual factors such as high rates of mobility, unpredictable security conditions, and competing economic demands often impede school access, teacher attendance, and student attendance

(Brown et al., 2023; Saavedra et al., 2019). For these reasons, among others, the dosage of the program—or the amount of intervention that the participants receive—tends to be lower in fragile contexts than in higher-income contexts, ultimately decreasing the quantity of instruction and practice opportunities, resulting in decreased learning outcomes.

Intervention dosage is a multidimensional construct, including program intensity (e.g., number of hours per day), frequency (e.g., days per week), and duration (e.g., number of months or years). This study focuses on program duration because of its significant policy implications for the field of education and interventions: providing longer duration of a program is a feasible method of increasing opportunity for students to attend the program and therefore increasing dosage—and impacts—of programming; on the other hand, longer program duration increases cost that may limit the number of children served by a program with a fixed budget. In general, increased duration of education intervention would likely correlate with increased learning outcomes (Shonkoff & Phillips, 2000). Beyond de facto increasing the treatment quantity of the program, longer duration correlates with increased efficiency. For example, implementation lags and kinks are likely to be worked out as a program matures and as program implementers benefit from some on-the-job experience (King & Behrman, 2009). Banerjee and colleagues (2007) report higher impacts with longer duration on two remedial education programs in India: (1) students who participated in a computer-adaptive math program for two years had higher improvement in math performance than students who participated for one year and (2) a small-group, pull-out remedial tutoring program that employed trained community members (“Balsakhi”) doubled its impact on students’ test scores after the program was implemented for two years, compared to the impact after the first year. While these results indicate that longer duration can result in higher impacts, there is little evidence on whether longer program duration *within* a school year is helpful, a question especially pertinent to humanitarian/crisis settings where program cycles lasting less than one academic year are common.

Despite some promising evidence associated with longer duration of programming, there are also several phenomena that can contribute to decreases in program impact with longer duration. A “pioneering effect” may result in early enthusiasm about a project that wanes over time; longer program duration provides additional time for treatment spillover to control groups; and many interventions may see positive impacts but at diminishing marginal returns, some of which may fall below minimally detectable effect sizes (King & Behrman, 2009).

Policy makers and program implementers must navigate the tension between serving the maximum number of students that resources will allow while ensuring the highest impact for those who have access to programming (Glewwe, 2013; Steuerle et al., 2007). Given the limited resources

available for educational programming in humanitarian contexts and the worldwide economic downturn, the study of program design factors, such as length of programming, is critical for determining the most effective and efficient methods of improving children’s educational outcomes (Wen et al., 2012).

### *SEL in Remedial Programming*

Despite increasing interest in social and emotional learning (SEL) from stakeholders working with refugees (GEMR, 2019), existing evidence on remedial programming often overlooks the potential of social and emotional practices in classrooms. A growing body of research, mostly from high-income countries, indicates that supporting SEL skills in classrooms either via inclusive teaching practices that target classroom climate or via more explicit, skill-targeted activities can contribute to increased academic outcomes and student well-being (Durlak et al., 2010, 2011). For example, climate-targeted SEL techniques may focus on positive classroom management and positive student-teacher relationships, whereas skill-targeted SEL activities, by contrast, provide explicit instruction on how to acquire and implement specific social and emotional skills and strategies, such as impulse control. Climate-targeted and skill-targeted SEL strategies can be utilized independently, though they are often implemented jointly (Jones & Bouffard, 2012).

*Healing Classrooms.* The evidence base on whether and how such SEL strategies work in non-Western contexts is limited. A small but growing number of trials indicates that the International Rescue Committee (IRC)’s Healing Classrooms curricula presents a promising programming strategy, particularly in emergency contexts.

*Version 1: Climate-targeted SEL.* Healing Classrooms is a classroom climate-targeted SEL program that provides teacher curricular materials and in-service training to deliver basic literacy and numeracy instruction in the context of safe and supportive formal school settings (Learning in a Healing Classroom [HCL]) or nonformal, after-school settings (Tutoring in a Healing Classroom [HCT]). When delivered via teachers in public schools in the Democratic Republic of the Congo, access to one year of HCL increased students’ perceptions of the supportiveness and predictability of teachers and schools while also improving literacy and numeracy skills (Aber et al., 2017; Torrente et al., 2015, 2019). It had no impact on children’s mental health problems or experiences of peer victimization (Torrente et al., 2015). When delivered via teachers as part of an after-school academic remedial program in a conflict-affected region of Niger, 22 weeks of access to HCT improved children’s French literacy and math skills compared to access to public school alone (Brown et al., 2023).

For HCT in Lebanon, each session began with a 10-minute introduction, followed by 40-minute blocks of instruction in Arabic, math, and second language (English or French) broken up by two 10-minute breaks and ended with a 10-minute wrap-up. Literacy instruction in both Arabic and the second language focused primarily on basic skills such as print awareness, letter recognition, phonemic awareness, and phonics that are directly teachable and support the acquisition of reading comprehension skills (Snow & Matthews, 2016). Math instruction focused on discrete instruction in number recognition, counting, place value, and addition and subtraction skills. All students received a school kit containing a workbook and school supplies to support students' ability to practice skills at home.

All teachers in the HCT program were Lebanese citizens recruited from local communities with, at minimum, a bachelor's degree. Teachers received a five-day preservice training on how to integrate five core classroom SEL principles—intellectually stimulating environment, sense of self-worth, sense of belonging, sense of self-control, and positive social relationships—into instruction through classroom management, critical thinking, and positive pedagogy. Teachers received an average of three follow-up visits per cycle from coaches who observed a 40-minute class period and provided feedback to teachers on cross-subject matter. Teachers in the HCT program also met in teacher learning circles (TLCs) once per month for 90 minutes, during which time they reflected upon, planned how to use, and practiced in a small group specific literacy and numeracy instructional and classroom management techniques. By combining climate-targeted SEL and instructional practices, we hypothesize that the HCT program would improve students' literacy and numeracy skills as well as support children's sense of safety and belonging in a schooling environment, as measured by their perceptions; a theory of change for HCT in Lebanon is depicted in Figure 1. The IRC estimates the unit cost of provision of the HCT package for two cycles—encompassing teacher professional development, tutoring, and school kit costs—was \$464 per child.

**Version 2: Climate-targeted SEL + skill-targeted activities:** In this version of HCT, each tutoring session followed the same schedule of instruction as the HCT-only classes, with one exception: the introduction and two subject-matter breaks were replaced by three 10-minute periods of targeted mindfulness activities (to improve stress reactivity and emotional/behavioral regulation) or brain games (to improve executive function) (90 minutes per week). For each session of a skill-targeted SEL activity, teachers were free to choose from a menu of activities that they deemed appropriate at the time of the session. These skill-targeted activities focused on strengthening specific social and emotional and cognitive-behavioral skills hypothesized to be impacted by exposure to trauma (Shonkoff et al., 2012) and critical for improving academic and SEL outcomes (Blair & Raver, 2012).

*Mindfulness activities* consist of brief mindfulness practices, such as breathing techniques and mindful movement. Focusing on mind-body “down-regulation” (Chiesa et al., 2013), these activities are designed to help children manage stress and regulate emotional responses and behaviors in the short-term, in the face of stressful/emotionally arousing events, and in the longer term—to improve general cognitive regulation (executive function), social and emotional functioning, and academic outcomes. *Brain games* consist of short games that use movement and play to practice executive function skills (attention, working memory, inhibitory control) necessary for learning (Jones et al., 2015a). Through the improvement of these core executive function skills, children are also better able to regulate their behavior and emotions in classrooms, leading to better academic gains (Finch et al., 2022; Kim et al., 2020; Suntheimer et al., 2022). We hypothesized that the HCT+Act program would help children better regulate stress, behavior, and cognitive and emotional reactions in social contexts compared to children with access to HCT; a theory of change of the HCT+Act is depicted in Figure 2.

All teachers in the HCT+Act program received two additional days of preservice training on mindfulness activities and two days on brain games activities that focused on how to select and integrate the activities into the HCT schedule and structure. In addition, teacher observations and learning circles incorporated support on the SEL activities, such as whether teachers used a calm tone of voice or selected activities appropriate to students' developmental age and needs. The international NGO estimates the unit cost of provision of the HCT+Act package for two cycles—encompassing the basic HCT costs but also the added TPD and SEL material costs—was \$531 per child.

The HCT+Act program was tested with a heterogeneous group of Nigerian refugees, Nigerian host community students, and Nigerian internally displaced students for 22 weeks during the 2016–17 academic year. Results from Niger indicated that one year of access to HCT+Skill-Targeted SEL programming improved French literacy and math skills *and* improved school grades compared to access to public school alone (Brown et al., 2023). The addition of skill-targeted SEL activities to HCT (HCT+Act) did not have a measurable impact on French literacy and math skills but did positively impact students' average school grades—an indicator of motivation and persistence as well as academic skill—compared to HCT alone. While HCT proved effective in raising academic outcomes, average student gains in absolute terms would not result in grade-level proficiency. In addition, the study was not able to collect data on a wide range of academic, social, and emotional outcomes.

#### *Previous Findings of HCT in Lebanon*

This study builds upon work by Tubbs Dolan et al. (2022) that leveraged a randomized waitlist-control group to test the

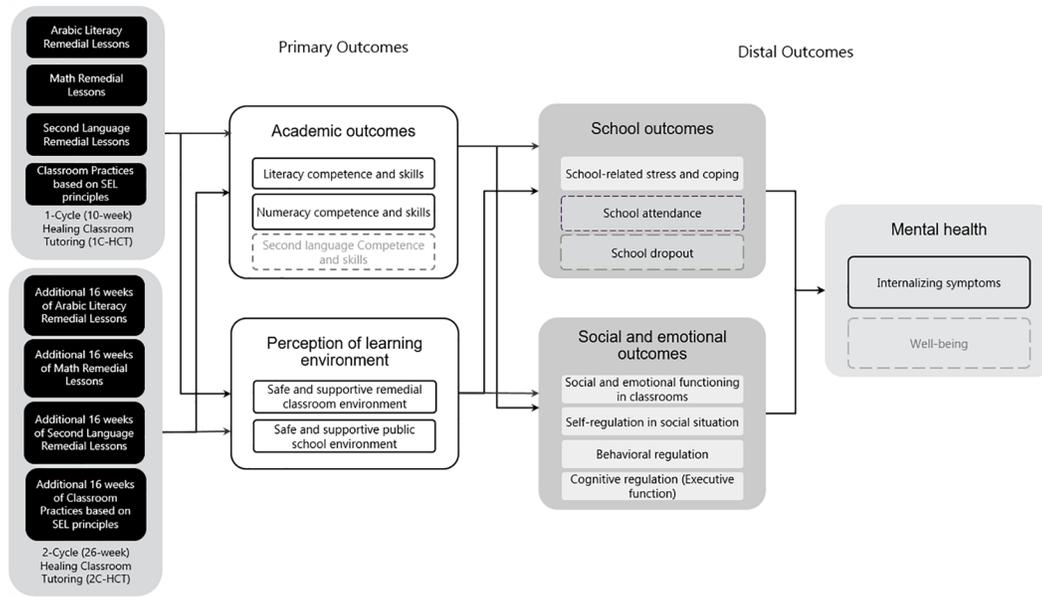


FIGURE 1. Theory of change for 1-cycle and 2-cycle Healing Classroom Tutoring programs.

impact of two versions of short-term HCT programming (climate-targeted Tutoring only [HCT] and the addition of skill-targeted SEL activities [HCT+Act]) on Syrian refugee children enrolled in Lebanese public schools. After 16 weeks of programming, students who had access to eight hours per week of HCT infused with climate-targeted SEL had improved perceptions of public school ( $ES = 0.48-0.66$ ) and behavioral regulation skills ( $ES = 0.24$ ) but no other measured academic or SEL outcomes. By contrast, students with access to the HCT+Act condition that implemented mindfulness-themed, skill-targeted SEL activities (above and beyond classroom-level SEL) performed better on discrete literacy and numeracy skills in addition to improved perceptions of public schools in comparison to peers who attended public school only ( $ES = .08-.12$ ). Contrary to expectation, children with access to HCT+Act reported significantly higher school-related stress at endline compared to those in the public-school-only condition, though both groups declined from baseline levels.

Similar to Niger, this first study in Lebanon had two important takeaways: (1) HCT+Act programming is a promising strategy for supporting academic and SEL outcomes among refugee children enrolled in public schools, particularly when compared to peers in public schools alone and (2) intervention effects, particularly on academic outcomes, were limited. Though there is potential in the programmatic approach, we hypothesize that iterating on its design may yield higher impacts.

### Present Study and Research Questions

In this study, we investigate whether we can improve upon the treatment(s) previously studied in Tubbs Dolan

et al.—either by increasing the duration of access to remedial tutoring or by enhancing the full-year treatment with additional SEL activities—to increase impacts on student academic or SEL outcomes. Targeted outcomes for the first contrast (increased duration) are student academic skills and perceptions of their public school and remedial classroom environments; targeted outcomes for the second (addition of skill-targeted SEL activities) are student perceptions of school-related stress, executive function skills, teacher-reported social-emotional functioning of students, and self-regulation. Specifically, we ask:

*RQ1.* Among Syrian refugee children enrolled in Lebanese public schools with access to 10 weeks of nonformal, classroom-climate SEL remedial programming, what is the impact of an additional 16 weeks of access to remedial programming on children’s academic and social outcomes?

*RQ2.* Among Syrian refugee children enrolled in Lebanese public schools with access to 26 weeks of nonformal, classroom-climate SEL remedial programming in Lebanon, what is the additional impact of 26 weeks of skill-targeted mindfulness and brain games activities?

## Methods

### Participants and Context

In SY2016–2017—the year in which the current study took place—there were an estimated 1,001,051 registered Syrian refugees in Lebanon, 55 percent of whom were under the age of 18 (UNHCR, 2017). As of January 2017, 195,706

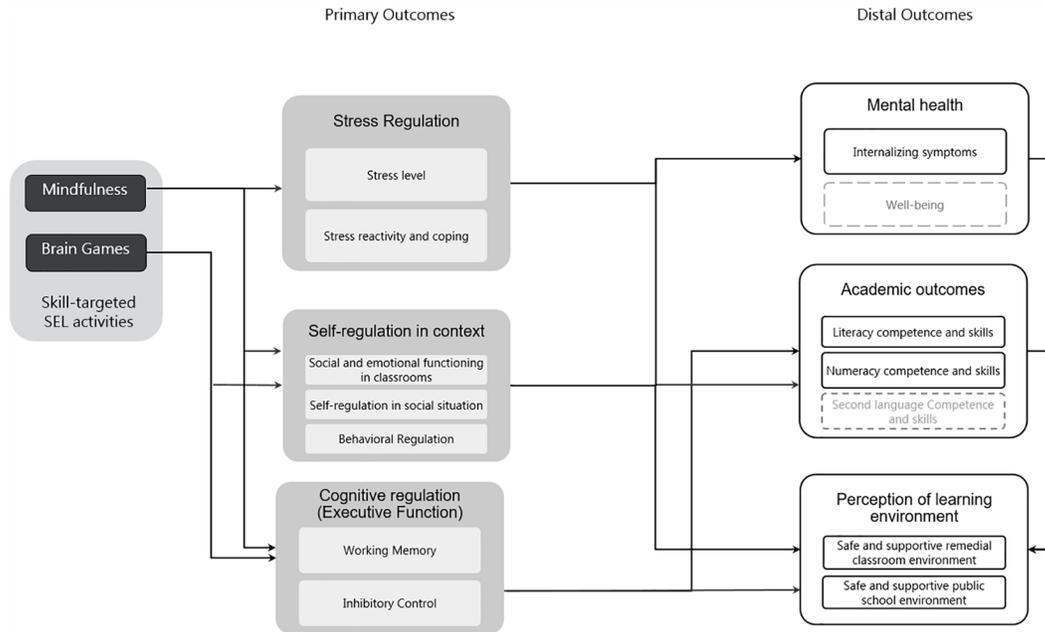


FIGURE 2. Theory of change for skill-targeted SEL activities.

Syrian children were enrolled in Lebanese public schools in first (71,566) and second (124,140) shifts (Abdul-Hamid & Yassine, 2020). All children in the current study attended public school in second-shift classes, which were held for four hours per day on weekdays, exclusively for Syrian children. Syrian students are intended to be placed in classes by ability (rather than age) according to an academic assessment, but practices at second shift were reportedly highly variable, as school staff adapted to local resource, political, and social needs and constraints (Adelman, 2018).

In 2016, the Lebanese Ministry of Education and Higher Education (MEHE) rolled out a nonformal education framework that outlined how NGOs could develop and deliver programs to support children in public schools. In 2016–2017, NGOs were permitted to offer noninstructional and supplementary services, including remedial classes—officially called retention support—to Syrian refugee students in Lebanese public schools (Buckner et al., 2018; Tubbs Dolan et al., 2022). In accordance with this policy, the IRC developed a remedial support program to support Syrian refugee students’ holistic learning outcomes in order to increase students’ ability to benefit from and persist in Lebanese public schooling.

All data presented here were collected as part of a large-scale, multiyear set of cluster-randomized controlled trials of nonformal, SEL-infused remedial programming in Lebanon. The current study focuses on evaluating the impact of the duration of remedial programming delivered over two consecutive program cycles (16 weeks and 10 weeks, respectively) during school year 2016–2017 to 76 community sites

recruited in the Akkar ( $j = 38$ ) and Bekaa ( $j = 38$ ) regions of Lebanon. Children registered within the first two weeks of the launch of the program and who had any record of student assessment data were included as part of the intent-to-treat sample ( $N = 4,017$ , 49% female; Figure 3). The final sample included students aged 5 to 15 ( $M = 8.98$ ,  $SD = 2.36$ ) attending grades 1 to 9 in Lebanese public schools ( $M = 2.76$ ,  $SD = 1.75$ ), with the vast majority of them attending grade six or lower (95%) and aged 12 or younger (91%). See Appendix A, Tables A1 and A2 for site and student demographic information by treatment condition.

### Design and Randomization

The 76 sites included in this study were stratified by region and randomized into one of the three treatment arms: (1) 33 public school + Tutoring in a Healing Classroom sites for the full year, in two consecutive program cycles, equivalent to 26 weeks (2C-HCT); (2) 33 public school + Tutoring in a Healing Classroom + skill-targeted SEL activities for the full year, in two consecutive program cycles (2 weeks) (2C-HCT+Act); and (3) 10 sites that were assigned to a waitlist control for the first program cycle of the year (16 weeks), during which they continued to attend public school but did not have access to remedial programming. This waitlist group then received access to HCT remedial programming for the second cycle of programming (10 weeks) (1C-HCT)<sup>1</sup>. The impacts of Cycle 1 programming are reported elsewhere (Tubbs Dolan et al., 2022). Here we report on the two pair-wise comparisons for the two

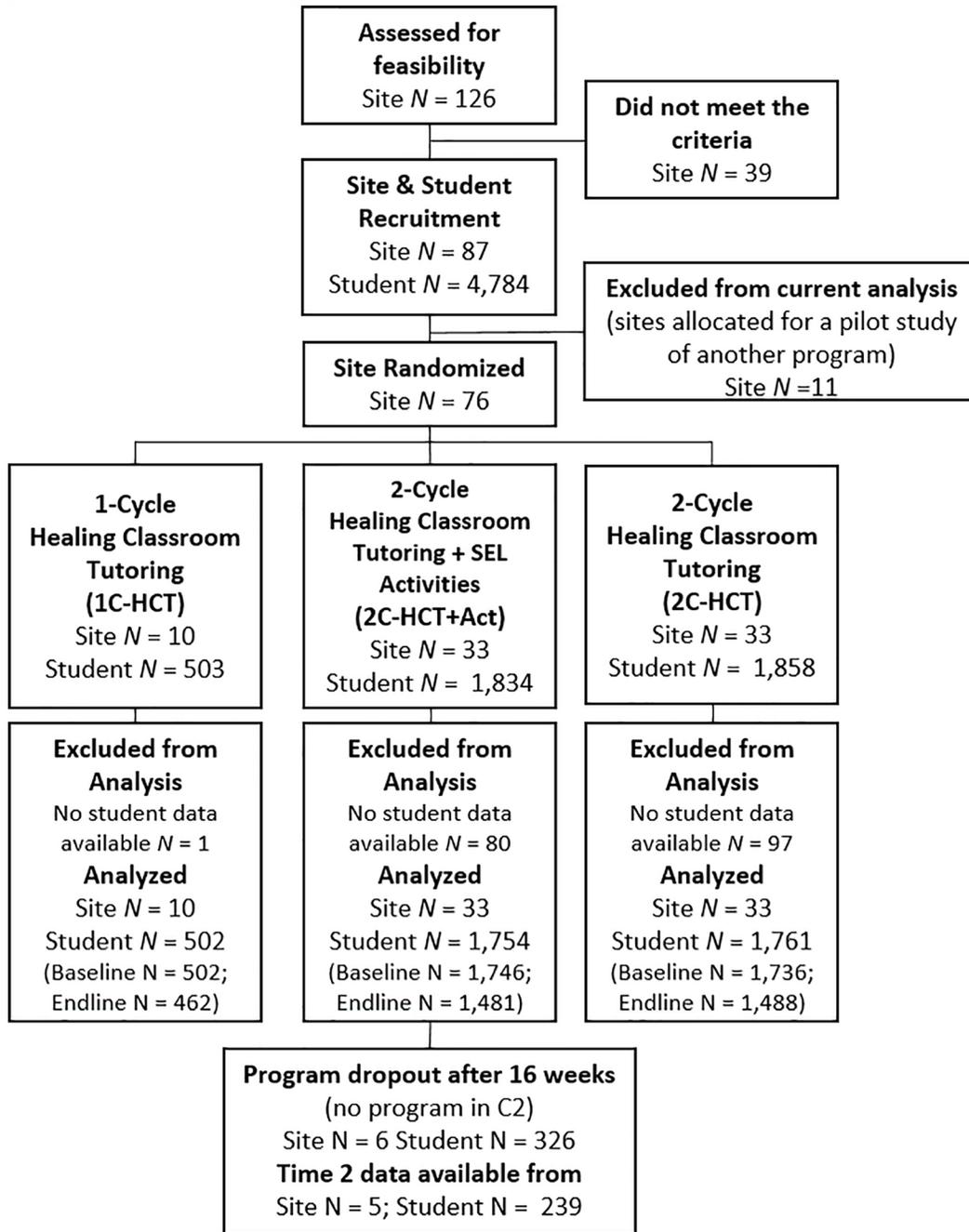


FIGURE 3. *Sampling and study design.*

treatment contrasts: (a) 33 2C-HCT sites vs. 10 1C-HCT sites and (b) 33 2C-HCT sites (same 2C-HCT sites as the first contrast vs. 33 2C-HCT+Act sites), corresponding to the two research questions.

Power analyses based on reasonable assumptions and accounting for the nested data structure of students within sites using Optimal Design software (Spybrook et al., 2006) suggested that the treatment contrast comparing one

versus two cycles of HCT (10 vs. 33, harmonic mean 15) had 80% power to detect the impact for a range of minimum effect sizes (MDESs) of 0.18–0.44 *SD*, when  $\alpha = 0.05$ ,  $ICC = 0.05$ – $0.20$ , and variance explained by covariates = 0.20–0.70, and the treatment contrast comparing two cycles of HCT and two cycles of HCT+Act 0.12 to 0.29 for the impact of two cycles of HCT versus two cycles of HCT+Act.

## Measures

Students were individually assessed through structured one-on-one verbal interviews in Arabic by trained local assessors prior to the intervention (baseline, November 2016) and again at the end of the program (endline, May to June 2017). A common set of priority measures (e.g., demographic characteristics, literacy skills) were assessed on the whole sample (core package). To minimize students' assessment burden while capturing a wide range of children's academic and social emotional functioning, some measures were assessed on a randomly selected half of the sample (Package A) and others on the other half of the sample (Package B). All measures used in this study were tested for and have demonstrated evidence of structural validity, internal consistency, correlational validity, and measurement invariance across the treatment groups, gender, and age, and across baseline and endline. Raw item statistics and psychometric details of all measures for this study are available in Gjicali et al. (2020a). A summary of measures is included in Table 1.

### Academic Outcomes

*Literacy skills and competence.* Students' Arabic literacy skills were measured using the Early Grade Reading Assessment (EGRA) (Gove & Wetterberg, 2011) adapted for the current study. EGRA included seven subtasks ranging from preliteracy to higher-order skills such as reading comprehension (see Table 5 for full list of subtasks). Due to floor effects and some ceiling effects, subtask scores that had more than nine items (all except reading comprehension and dictation) were transformed into proficiency levels with zero scores coded as zero and nonzero scores coded as 1–5 based on percentile ranks (20, 40, 60, 80, 100) based on the baseline distribution. A latent factor consisting of all subtask scores was used to represent children's overall literacy competence. In addition, subtask scores were used to test impacts on discrete literacy skills.

*Numeracy skills and competence.* The Early Grade Mathematics Assessment (EGMA) (Dubeck & Gove, 2015) was used to capture number skills and competence (see Table 5 for subtasks). Subtask scores containing more than nine items were transformed using the same quintile transformation strategy used for EGRA (except word problems, which had six items and ranged 0–6). A latent factor, identified through factor analysis, consisting of all recoded subtask scores was used to represent children's overall numeracy competence, and the subtask scores were used to test impacts on observed discrete numeracy skills.

### Social and Emotional Outcomes

*Perception of public school environment.* Select items from the Child Friendly School Questionnaire (CFSQ) were used to measure students' perceptions of their public school

environment (Godfrey et al., 2012; see Gjicali et al., 2020b, for psychometric details).

*Perception of remedial classroom environment.* Of the CFSQ items used to capture children's perceptions of their public school environment, we selected 13 items to capture students' experience in tutoring programs, measured at endline only. Exploratory and confirmatory factor analyses suggested that the items consist of two constructs representing: (1) positive climate (10 items, e.g., The IRC Remedial School is a welcoming and inviting place for families like mine; I look forward to coming to the IRC Remedial School); and (2) engaging and safe remedial school (3 items, e.g., I sometimes stay home from the IRC Remedial School because I am worried about my safety [reverse coded]).

*Teacher reports of social and emotional functioning.* We used items from the Strength and Difficulties Questionnaire (SDQ: Goodman et al., 2000) and Classroom Executive Function Survey (CEFS: Jones et al., 2015b) to capture teachers' assessment of children's social-emotional functioning in classrooms. Because the children assigned to the IC-HCT group did not have access to remedial tutoring in cycle one, we do not have baseline data on children in this group.

SDQ is a behavioral screening instrument for children, consisting of five subscales: Hyperactivity, Emotional Symptoms, Conduct Problems, Peer Problems, and Prosocial Behavior. CEFS captures teacher perception of students' executive function skills. Due to inadequate model fit of the original SDQ subscales and overlapping constructs across two measures, we combined all teacher-report items and generated five subscales based on exploratory and confirmatory factor analyses using a total of 29 items (see details in Gjicali et al., 2020a).

*School-related stress and stress reactivity.* Two subscales of the Response to Stress Questionnaire-Academic Problems (RSQ-AP: Connor-Smith et al., 2000) were used to capture Syrian refugee children's school-related stress and stress reactivity (Package A). Specifically, the Academic Problems subscale was used to measure the perceived level of stress in school-related academic events and the Response to Stress: Involuntary Engagement subscale was used to capture children's involuntary stress reactivity to their perceived school-related stress. These subscales have shown evidence of reliability and validity with Syrian refugee children in Lebanon (Kim et al., 2021).

*Behavioral regulation.* For the full sample, children's behavioral regulation was rated by assessors using an adapted version of the Preschool Self-Regulation Assessment-Assessor Report (PSRA-AR: McCoy et al., 2017; Smith-Donald et al., 2007). Assessors rated each child on the

TABLE 1  
Summary Measure Table

Construct	Measure	Report	Packet	Items	Range	Cronbach's $\alpha$
<u>Academic Outcomes</u>		PM				
Arabic literacy skills	Early Grade Reading Assessment (EGRA)		Core	6 subtasks	0–5	0.92–0.93
Math skills	Early Grade Math Assessment (EGMA)		A	9 subtasks	0–6	0.95–0.96
<u>Perceptions of Public School Environment</u>		CR	B	19	1–4	
Caring and supportive teachers	Child-Friendly School Questionnaire			4		0.63–0.92
Engaging and motivating school	Child-Friendly School Questionnaire			7		0.72–0.82
Respectful and inclusive school	Child-Friendly School Questionnaire			5		0.81–0.92
Safe school	Child-Friendly School Questionnaire			3		0.64–0.65
<u>Perceptions of Remedial Classroom Environment</u>		CR	B	13	1–4	
Positive climate	Child-Friendly School Questionnaire			10		0.90
Engaging and safe remedial school	Child-Friendly School Questionnaire			3		0.66
<u>Teacher Report of Social &amp; Emotional Functioning</u>		TR	Core	29	1–4	
Pro-social behavior and social acceptance	Factors from combined Strength and Difficulties Questionnaire & Classroom Executive Function Survey (see measures section for details)			9		0.85–0.87
Hyperactive and externalizing behaviors				6		0.77–0.77
Emotional distress and internalizing behaviors				6		0.66–0.67
Working memory functioning				4		0.88–0.91
Inhibitory control functioning				4		0.85–0.89
<u>School-Related Stress and Stress Reactivity</u>		CR	A	25	1–4	
School-related stress	Response to Stress Questionnaire: Academic Problems			10		0.89–0.92
Involuntary stress reactivity	Response to Stress: Involuntary Engagement			15		0.91–0.92
<u>Behavioral Regulation</u>		AR	Core	13	1–5	0.96–0.95
<u>Cognitive and Emotional Regulation</u>		HS	B	24		
Hostile attribution bias	Children's Stories, adapted			6	0–1	0.79–0.85
Anger dysregulation	Children's Stories, adapted			6	1–5	0.89–0.92
Sadness dysregulation	Children's Stories, adapted			6	1–5	0.89–0.92
Aggressive reaction	Children's Stories, adapted			6	1–3	0.80–0.73
<u>Internalizing Symptoms</u>		CR	Core	13	0–2	0.96–0.95
<u>Executive Function</u>		PM	A			
Working memory	Rapid Assessment of Cognitive and Emotional Regulation			n/a	–5.78 –1.79	n/a
Inhibitory control	Rapid Assessment of Cognitive and Emotional Regulation			n/a	–8.74 –9.12	n/a

Note. CR = child report; TR = teacher report; AR = assessor report; HS = hypothetical scenarios; PM = performance measure.

behaviors displayed. Psychometric details and evidence of reliability and validity with the current sample is presented in Wu et al. (2020).

*Cognitive and emotional regulation.* Children's Stories (CS: Dodge et al., 2015) was administered to measure children's cognitive and emotional regulation. CS is a scenario-based assessment designed to measure hostile attribution bias and reactive aggression and has been previously validated in six countries, including Jordan (Dodge et al., 2015). We adapted CS to include six contextually appropriate, hypothetical scenarios of ambiguous peer interactions that could lead to social conflict. Each scenario was followed by a series of questions to assess children's hostile attribution bias and self-predicted reactive aggression. In addition, we added two questions about emotional dysregulation (sadness, anger) taken from the emotion regulation measure used by Di Giunta et al. (2017). See details of the subscales and psychometric evidence in Gjicali et al. (2020a).

*Internalizing symptoms.* The Arabic version of the Moods and Feelings Questionnaire (MFQ: Tavitian et al., 2014) was used to measure internalizing. MFQ asks a child to report the frequency of experiencing certain feelings or behaviors in the past two. The MFQ has been previously validated for 5- to 15-year-olds in Lebanon (Tavitian et al., 2014).

*Executive function: working memory, inhibitory control.* The Rapid Assessment of Cognitive and Emotional Regulation (RACER: Ford et al., 2019) was used to assess two aspects of executive function: working memory and inhibitory control. Working memory was measured using a Spatial Delayed Match to Sample task (Goldman-Rakic, 1996) and inhibitory control was measured using a Simon Task (Simon & Rudell, 1967). Children were asked to play tablet-based games containing these tasks, and scores for each task were obtained from their performance in each game. For task and scoring details, see Ford et al. (2019). For ease of interpretation, the working memory and inhibitory control scores were standardized using baseline mean and standard deviation.

*Covariates.* Information on child- and site-level characteristics were used as covariates in the impact models. From administrative records, we retrieved information on site characteristics, child demographic characteristics, and child screening test scores on Arabic reading, math, and second language (Annual Status of Education Report [ASER]: Banerji et al., 2013). In addition, through structured parent interviews we collected baseline information on potential risk factors common among the Syrian refugee population, various indicators of socioeconomic status, and other household and child characteristics. See Appendix A for

descriptive information on all covariates used in the impact analyses.

### *Recruitment and Implementation*

The IRC recruited research sites and students through 167 community awareness sessions in August and September 2016. In accordance with MEHE's Nonformal Education framework, the IRC identified students in each community targeted for the remedial program: students aged 5 to 15 who were enrolled in local public schools. The IRC actively identified and registered children until the end of November, two weeks after the program launch. Parents and guardians were informed about research activities and were asked to provide informed consent. A site and participant recruitment and randomization process is presented in Figure 3.

In the first program cycle (November–March), the planned dosage for each of the 66 sites was 48 tutoring sessions held over 16 weeks, in addition to one week of screening and orientation sessions (up to three additional sessions). On average, programming sites in the first programming cycle reported completing the intended dosage (2C-HCT:  $M = 48.84$ ;  $SD = 1.72$ ; 2C-HCT+Act:  $M = 48.82$ ;  $SD = 1.72$ ). In the second programming cycle (March–June), 30 sessions were intended for the evaluation period over the course of 10 weeks for all participating sites; the evaluation period was shorter in the second programming cycle due to anticipation of low attendance after 10 weeks of programming because of Ramadan, seasonal migration patterns, and the end of the public school year. On average, 1C-HCT and 2C-HCT sites reported completing the intended dosage (1C-HCT:  $M = 29.19$ ;  $SD = 0.96$ ; 2C-HCT:  $M = 29.20$ ;  $SD = 0.97$ ). However, due to the implementing partner's decision to cease programming in six 2C-HCT+Act sites in the second program cycle, this group had lower dosage overall,  $M = 24.97$ ;  $SD = 11.52$ ). The 2C-HCT+Act sites that actively implemented programming did so at full intended dosage on average,  $M = 29.47$ ,  $SD = 0.80$ . Across both cycles, students attended about half or less than half of the remedial sessions offered (1C-HCT:  $M = 51\%$ ;  $SD = 35\%$ ; 2C-HCT:  $M = 42\%$ ;  $SD = 31\%$ ; 2C-HCT+Act:  $M = 43\%$ ;  $SD = 32\%$ ).

### *Data Collection*

Baseline data collection occurred in the regions of Akkar and Bekaa over three weeks in mid-November to early December 2016; endline data was collected following the same procedures in three weeks from mid-May to the first week of June 2017. Data collectors conducted one-on-one verbal assessments in a quiet, but not secluded, area near tutoring classes (for children attending tutoring) or at children's homes (for children not attending tutoring at baseline or who were absent on assessment days). Performance-based measures, including EGRA, EGMA, and RACER, were

TABLE 2

*Unadjusted Baseline Equivalency: Treatment Group Mean Differences in Study Outcomes by Contrast*

	Full-year Healing Classrooms Tutoring vs. Half-year Healing Classrooms Tutoring (2C-HCT – 1C-HCT)				Full-year Healing Classrooms Tutoring +SEL Activities vs. Full-Year Healing Classrooms Tutoring (2C-HCT+Act – 2C-HCT)			
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>q</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>q</i>
Perceptions of the Public School Environment								
Caring and supportive teachers	0.102	0.094	0.277	1.000	0.040	0.114	0.724	1.000
Engaging and motivating school	0.026	0.080	0.743	1.000	-0.103	0.087	0.238	1.000
Respectful and inclusive school	0.148	0.119	0.216	1.000	-0.068	0.167	0.682	1.000
Safe school	0.154	0.116	0.184	1.000	-0.163	0.131	0.213	1.000
Teacher-report Social and Emotional Functioning <sup>a</sup>								
Prosocial behavior and social acceptance	–	–	–	–	-0.072	0.151	0.635	1.000
Hyperactive and externalizing behaviors	–	–	–	–	0.115	0.103	0.262	1.000
Emotional distress and internalizing behaviors	–	–	–	–	0.272	0.142	0.055	1.000
Working memory functioning	–	–	–	–	-0.038	0.225	0.867	1.000
Inhibitory control functioning	–	–	–	–	-0.007	0.094	0.941	1.000
Academic Outcomes								
EGRA literacy competence	-0.155	0.152	0.310	1.000	0.116	0.086	0.185	1.000
EGMA numeracy competence	-0.097	0.159	0.540	1.000	0.070	0.088	0.427	1.000
School-Related Stress and Stress Reactivity								
School-related stress	-0.223	0.101	0.027	0.276	0.024	0.100	0.813	1.000
Stress reactivity	-0.175	0.076	0.022	0.276	0.031	0.059	0.599	1.000
Behavioral regulation	-0.038	0.099	0.697	1.000	0.136	0.245	0.578	1.000
Cognitive and Emotional Regulation								
Hostile attribution bias	-0.090	0.081	0.266	1.000	0.072	0.053	0.169	1.000
Anger dysregulation	0.089	0.106	0.399	1.000	0.111	0.112	0.321	1.000
Sadness dysregulation	-0.104	0.121	0.389	1.000	0.028	0.142	0.843	1.000
Aggressive reaction	0.033	0.115	0.775	1.000	0.155	0.144	0.282	1.000
Internalizing symptoms	-0.035	0.098	0.721	1.000	0.028	0.053	0.595	1.000
Executive Function <sup>b</sup>								
Working memory	-0.035	0.083	0.676	1.000	-0.045	0.054	0.413	1.000
Inhibitory control	-0.046	0.086	0.595	1.000	0.173	0.098	0.081	1.000

*Note.* Baseline equivalency estimates are estimated from the mean differences in the treatment-group measurement invariances models of each measure, standardized based on the reference group mean and variance. Baseline equivalency for the Perceptions of HCT Classroom Environment is unavailable because it was not assessed at baseline.

<sup>a</sup>Teacher Report Social and Emotional Functioning was not administered to the control group (1C-HCT) at baseline and, therefore, is not available for the dosage contrast.

<sup>b</sup>For Executive Function, measurement model is not available; instead, we present the standardized coefficients of treatment indicators.

conducted using tablets. All other measures were collected via paper-and-pencil forms. Each child assessment lasted about 45 minutes to an hour. Each day, data collectors digitized the paper-and-pencil survey forms and uploaded the data collected via tablets. Children's demographic and family characteristics were collected at baseline through structured interviews with parents or guardians.

#### *Baseline Equivalence Across Treatment Conditions*

Table 2 presents differences in child outcomes between treatment conditions at baseline, for each treatment contrast (RQ1: 2C-HCT vs. 1C-HCT; RQ2: 2C-HCT+Act vs. 2C-HCT). Results are from the scalar measurement invariance model across treatment groups, where the control condition in each contrast is fixed at a mean of 0 and variance of

1. We chose this model-based scoring approach to estimate and compare students' scores to maximize precision and power for impact estimation by minimizing measurement error, in recognition of the lack of normed and validated measures in the context and population. Without contextually valid norms and adequate comparisons with other populations, summary scores would not be prudent to interpret. It is worth noting that while we rigorously tested the psychometric properties of these measures, they are only interpretable when compared within the population currently presented in this paper. Based on this scoring and modeling approach, we found no significant differences across treatment conditions or measures after FDR correction for multiple outcomes.

### *Analytic Strategy*

All descriptive analyses were conducted using Stata SE version 15.1, and measurement and impact modeling was conducted using Mplus 8.3 (Muthén & Muthén, 2018), with mean- and variance-adjusted weighted least squares (WLSMV) estimator to address the non-normally distributed or categorical variables in the model. WLSMV estimation is based on the covariance matrix of all available information from the full sample and then fits the model with pairwise present data. Therefore, all available information is used from all cases to preserve the full sample with the WLSMV estimator, which has been empirically shown to produce consistent estimates under various missing data assumptions (Asparouhov & Muthén, 2010). To account for children nested within different research sites, cluster-robust standard errors were estimated using a sandwich estimator. This approach is an effective and efficient way to model complex data when sample size at the cluster level is not small (Huang, 2016). All models were evaluated using the conventional model fit criteria: an upper limit of 0.05 for the RMSEA and SRMR statistics (Kline, 2011) and a lower limit of 0.95 for the CFI and TLI (Hu & Bentler, 1999). Analyses testing treatment impacts were conducted using a series of structural equation models (SEM) estimated separately for each measure. For example, four outcomes included in CFSQ were modeled in a single structural model (see Figure 4). The measurement models for baseline and endline for each outcome were held scalar-invariant, with baseline factor scaled with mean of 0 and variance of 1 and included in the impact analysis SEM models to test baseline-adjusted effects of treatment and over and above the longitudinal change of the outcome in the control group. SEM can facilitate the practical interpretation of impact by placing effect size estimates on the same scale as within-individual growth over time.

A series of impact models were run for each outcome measure to answer each of the research questions corresponding to the two treatment contrasts. For each research

question, we a priori specified a set of primary (targeted) and secondary (exploratory) outcomes based on the intervention descriptions, prior evidence, and theory (see Present Study and Research Questions). We estimated the impacts separately for each outcome measure to account for covariance between conceptual constructs within each measure. For example, all four constructs of CFSQ were modeled in a single structural model (see Figure 4). This latent modeling analytic approach was used on all outcome measures, with the exception of the observed outcome variables (RACER working memory and inhibitory control; EGRA literacy and EGMA numeracy subtask scores).

To address research question one, the differences between 26-week HCT and 10-week HCT groups in student outcomes were estimated by including a dummy-coded 26-week HCT treatment condition indicator as a predictor in addition to the same outcome-latent factor measured at baseline (see Figure 4); similarly for research question two, the differences between 2C-HCT and 2C-HCT+Act groups were estimated using a 2C-HCT+Act indicator. The coefficients of each of the treatment groups can be interpreted as the impact of access to programming over and above the residual change between time points in the targeted outcome among children in the control group in each treatment contrast.

Both unadjusted and adjusted impact models were estimated for the impact analyses. We interpret the treatment impact estimates adjusted for various site and child covariates as the final impact estimates, as these account for residual error variance in the student outcomes for more sensitive impact estimate detection and to adjust for potential baseline nonequivalence across the treatment groups (see Appendix A1 and A2). In addition, given the site-level dropouts in 2C-HCT+Act condition, a sensitivity analysis excluding the six dropout sites was conducted to examine site-level treatment-on-the-treated impacts of the 2C-HCT+Act program ( $j = 27$ ) when compared to 2C-HCT condition ( $j = 33$ ) to test the robustness of RQ2.

All baseline latent variable measurement models were scaled to have a variance of 1 and mean of 0, and unstandardized impact estimates are reported as effect sizes scaled based on the variance of the baseline outcomes of the pooled sample. Similarly, the RACER score was standardized based on baseline mean and variance. The  $p$ -values for the final impact estimates for the primary and secondary outcomes of each treatment contrast were adjusted for type I error common in multiple hypothesis testing by controlling the sharpened false discovery rate (FDR: Benjamini et al., 2006). Both naive  $p$ -values and FDR-adjusted  $q$ -values are reported. Results significant at  $q < .05$  are interpreted as statistically significant findings; results significant at  $q < .10$  are interpreted as marginally statistically significant findings. We also report findings based on significant ( $<.05$ ) and marginal ( $<.10$ )  $p$ -values for hypotheses-generating and program revision

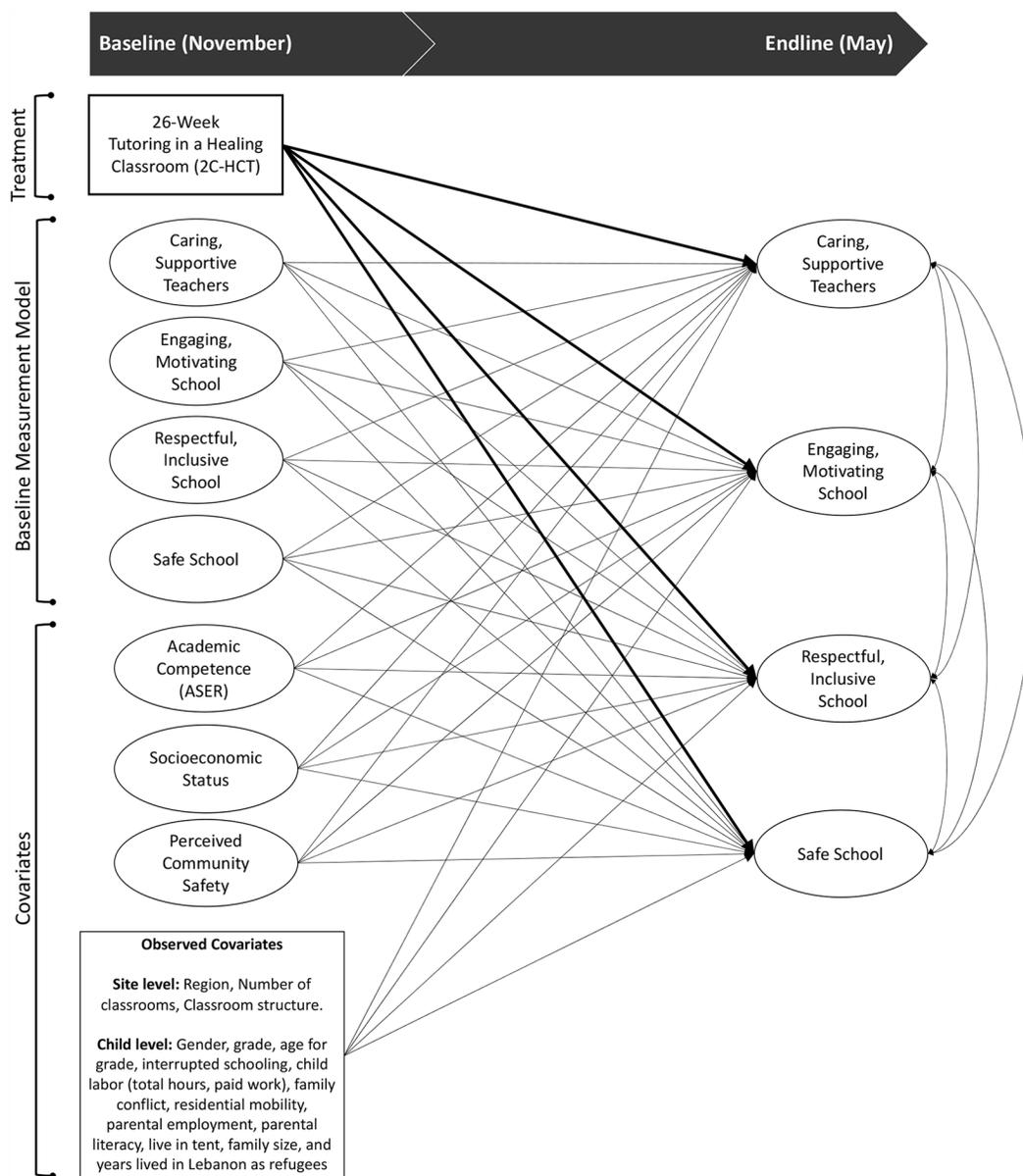


FIGURE 4. Simplified conceptual model of the final models testing the impact of 2C-HCT program (26 weeks), compared to 1C-HCT (10 weeks), for Perceptions of Public-School Environment outcomes as an example. Residual variances of all latent factors and covariances among exogenous variables were omitted for brevity.

purposes, given the small treatment contrasts and scant evidence of educational program evaluation in humanitarian contexts in low- and middle-income countries.

## Results

### Attrition and Missing Data

Figure 3 provides a flow chart of sample and attrition. Of the 76 intent-to-treat sites initially randomized, six sites

(student  $n = 326$ ) from the 2C-HCT+Act sites dropped out after Cycle 1 due to budgetary limitations.<sup>2</sup> Despite the discontinuation of programming in these six sites after Cycle 1, endline data were collected from the majority of the students in five of these dropout sites ( $n = 239$ , 73%); one site was inaccessible due to security concerns.

At the student level, inclusion in the ITT and analytic sample is defined as children who registered during the first two weeks of the programming and had any child assessment

TABLE 3

*Unadjusted Change Estimates Between Baseline and Endline in Primary Study Outcomes*

	Intercept Change for 1 Cycle HCT-Only Group				Intercept Change for HCT-Only Group (2 Cycles of HCT Basic)			
	<i>b</i>	<i>SE</i>	<i>p</i>	<i>q</i>	<i>b</i>	<i>SE</i>	<i>p</i>	<i>q</i>
Perceptions of Public School Environment								
Caring and supportive teachers	1.936	0.196	0.000	.001	0.952	0.211	0	0.001
Engaging and motivating school	1.305	0.155	0.000	.001	0.423	0.138	0.002	0.002
Respectful and inclusive school	2.010	0.209	0.000	.001	1.038	0.226	0	0.001
Safe school	0.379	0.121	0.002	.002	0.092	0.095	0.335	0.115
School-Related Stress and Stress Reactivity								
School-related stress	-0.224	0.120	0.063	.014	-0.25	0.084	0.003	0.002
Stress reactivity	-0.494	0.143	0.001	.001	-0.429	0.119	0	0.001
Cognitive and Emotional Regulation								
Hostile attribution bias	-0.555	0.122	-4.558	0	-0.535	0.083	0	0.001
Anger dysregulation	-0.519	0.097	-5.358	0	-0.441	0.119	0	0.001
Sadness dysregulation	-0.418	0.129	-3.251	0.001	-0.473	0.102	0	0.001
Aggressive reaction	-0.102	0.1	-1.02	0.308	-0.041	0.092	0.657	0.213
Behavioral regulation	0.240	0.090	0.008	.004	0.333	0.079	0	0.001
Internalizing symptoms	-0.755	0.114	0.000	.001	-0.64	0.104	0	0.001
Academic Outcomes								
EGRA literacy competence		0.019	0.000	.001	0.216	0.037	0	0.001
EGMA numeracy competence	0.331	0.038	0.000	.001	0.222	0.01	0	0.001
Executive Function <sup>a</sup>								
Working memory	0.097	0.033	0.003	.002	0.022	0.036	0.549	0.187
Inhibitory control	0.060	0.049	0.220	.047	-0.01	0.036	0.779	0.242

*Note.* The change estimates are from the unadjusted impact model, controlling for a treatment indicator, and represent change rate of each outcome from baseline to endline in the absence of treatment. Perceptions of Remedial Classroom Environment and Teacher-report Social and Emotional Functioning has no baseline data in some or all of the treatment groups and, therefore, no growth estimate is available.

data, either at baseline or endline. This resulted in the exclusion of 178 students (4%) who registered but did not have any record of attendance *or* child assessment data. Given the challenges of student-level tracking, high population mobility, and the programmatic eligibility requirement that students must be enrolled in and attending Lebanese public school, it is unclear whether this attrition was due to noncompliance, record-keeping problems (false record of registration),<sup>3</sup> or the result of children losing access to public schools (and thus program eligibility). Of the students included in the final sample, 32 (0.8%) were present in baseline data collection but missing in endline (1C-HCT  $n = 0$ ; 2C-HCT  $n = 24$ ; 2C-HCT+Act:  $n = 8$ ), and 585 (15%) were missing in baseline but present in endline (1C-HCT  $n = 40$  (8%); 2C-HCT  $n = 272$  (15%); 2C-HCT+Act:  $n = 273$  (15%)); one student lacked data at both time points with only an administrative record available. All available data from all 76 sites and 4,017 children included in the ITT sample were retained in the analysis, regardless of the program dropout and missing data in assessments, by addressing missing data through the modeling with WLSMV estimation.

#### *Change in Outcomes of Syrian Refugee Children in HCT Conditions*

The unadjusted change in all study outcomes between baseline and endline among children in the 1C- and 2C-HCT groups are provided in Table 3 to provide reference for change in the counterfactual as well as to provide pre- and post-test difference in these groups.<sup>4</sup> Overall, children in both conditions had significant change in the majority of the tested outcomes, all in positive directions. The only outcomes that did not show significant improvement are aggressive reactions in social conflict situations in both conditions and perceptions of safe school environment, working memory, and inhibitory control skills in the 2C-HCT groups.

#### *RQ1: Impacts of Additional Duration of Tutoring in a Healing Classroom (10 vs. 26 Weeks of HCT)*

*Impacts on Primary Outcomes.* First, we found that Syrian refugee children who had access to two cycles (26 weeks) of HCT reported significantly less positive perceptions of their public school environment as compared to their

TABLE 4

Impact Effect Size (ES) Estimates of 26 Weeks of Tutoring in a Healing Classroom (2C-HCT) Over and Above 10 Weeks (1C-HCT)

	Unadjusted				Adjusted			
	ES	SE	P	q	ES	SE	p	q
<b>Primary Outcomes</b>								
Perceptions of Public-School Environment								
Caring and supportive teachers	-0.984	0.217	0.000	0.001	-0.892	0.259	0.001	0.003
Engaging and motivating school	-0.883	0.159	0.000	0.001	-0.826	0.185	0.000	0.001
Respectful and inclusive school	-0.972	0.208	0.000	0.001	-0.867	0.276	0.002	0.003
Safe school	-0.287	0.116	0.014	0.012	-0.240	0.158	0.128	0.059
Perceptions of Remedial Classroom Environment								
Positive climate	-0.247	0.059	0.000	0.001	-0.243	0.065	0.000	0.001
Engaging and safe remedial school	-0.134	0.057	0.019	0.013	-0.149	0.064	0.020	0.017
Academic Outcomes								
EGRA literacy competence <sup>a</sup>	0.002	0.022	0.914	0.129	0.041	0.024	0.097	0.051
EGMA numeracy competence	-0.115	0.053	0.031	0.019	0.031	0.058	0.588	0.172
<b>Secondary Outcomes</b>								
Teacher-Report Social and Emotional Functioning								
Prosocial behavior and social acceptance	-0.052	0.077	0.505	1.000	0.083	0.087	0.340	1.000
Hyperactive and externalizing behaviors	-0.001	0.059	0.991	1.000	-0.025	0.049	0.611	1.000
Emotional distress and internalizing behaviors	0.079	0.073	0.281	1.000	0.134	0.073	0.065	0.835
Working memory functioning	-0.048	0.061	0.432	1.000	0.083	0.056	0.136	1.000
Inhibitory control functioning	-0.031	0.071	0.656	1.000	0.065	0.058	0.262	1.000
School-Related Stress and Stress Reactivity								
School-related stress	-0.026	0.129	0.839	1.000	0.090	0.103	0.383	1.000
Stress reactivity	0.065	0.150	0.663	1.000	0.136	0.117	0.244	1.000
Behavioral regulation	0.093	0.092	0.316	1.000	0.308	0.088	0.000	0.001
Cognitive and Emotional Regulation								
Hostile attribution bias	0.007	0.047	0.887	1.000	0.040	0.162	0.808	1.000
Anger dysregulation	0.030	0.038	0.429	1.000	0.027	0.119	0.822	1.000
Sadness dysregulation	-0.022	0.048	0.651	1.000	-0.039	0.127	0.759	1.000
Aggressive reaction	0.032	0.042	0.439	1.000	-0.009	0.093	0.923	1.000
Internalizing symptoms	0.115	0.120	0.340	1.000	0.086	0.125	0.492	1.000
Executive Function								
Working memory	-0.075	0.047	0.111	1.000	0.033	0.119	0.782	1.000
Inhibitory control	-0.070	0.060	0.239	1.000	0.037	0.920	0.689	1.000

counterparts with access to only one cycle (10 weeks). Specifically, they reported less caring and supportive public school teachers (adjusted  $ES = -0.89, p < .001, q = .003$ ); less engaging and motivating public school climate (adjusted  $ES = -0.83, p = .000, q = .001$ ); and less respectful and inclusive public schools (adjusted  $ES = -0.87, p = .002, q = .003$ ) (see Table 4). We additionally found that Syrian refugee children with access to two cycles of HCT reported significantly less positive perceptions of their remedial site, as compared to one cycle. Specifically, they reported perceptions of less positive climate (adjusted  $ES = -0.24, p < .001, q = .001$ ) and less

engaging and safe remedial site environment (adjusted  $ES = -0.15, p = .020, q = .017$ ) compared to children with access to one cycle.

Second, findings from the academic outcome models suggested that access to two cycles (26 weeks) of HCT had marginally positive impacts on overall literacy competence (adjusted  $ES = 0.04, p = .097, q = .051$ ) and no impacts on numeracy when compared to students with access to 10 weeks (Table 5). However, analyses of observed EGRA and EGMA subtasks scores suggested that access to 26 weeks did not have any detectable effects on discrete numeracy and literacy skills as compared to 10 weeks (Table 5).

TABLE 5

*Impact Estimates for the 26 Weeks of Tutoring in a Healing Classroom (2C-HCT) Over and Above 10 weeks (1C-HCT) on Literacy and Numeracy Subtasks*

	Unadjusted				Adjusted			
	$\beta$	<i>SE</i>	<i>p</i>	<i>q</i>	$\beta$	<i>SE</i>	<i>p</i>	<i>q</i>
EGRA Literacy Skills (Subtasks)								
Receptive vocabulary	0.064	0.065	0.327	1.000	0.039	0.044	0.368	0.290
Letter recognition	-0.029	0.026	0.264	1.000	0.032	0.035	0.352	0.290
Grapheme	-0.033	0.021	0.107	1.000	0.026	0.030	0.393	0.290
Invented word reading	-0.042	0.032	0.183	1.000	0.038	0.038	0.325	0.290
Oral passage reading	0.001	0.021	0.952	1.000	0.051	0.029	0.080	0.173
Reading comprehension	0.002	0.026	0.936	1.000	0.065	0.032	0.042	0.145
Word dictation	0.006	0.029	0.828	1.000	0.073	0.031	0.017	0.136
EGMA Numeracy Skills (Subtasks)								
Number identification	-0.046	0.040	0.253	0.409	0.036	0.044	0.412	0.926
Number discrimination	-0.089	0.032	0.005	0.048	-0.007	0.045	0.871	1.000
Missing number	-0.060	0.040	0.129	0.292	0.008	0.037	0.835	1.000
Addition level 1	-0.038	0.034	0.271	0.409	0.041	0.043	0.339	0.904
Addition level 2	-0.022	0.032	0.482	0.432	0.074	0.036	0.037	0.200
Subtraction level 1	-0.022	0.027	0.425	0.432	0.054	0.039	0.168	0.645
Subtraction level 2	-0.005	0.031	0.881	0.685	0.092	0.041	0.023	0.200
Multiplication	-0.071	0.043	0.098	0.292	0.048	0.050	0.335	0.904
Word problems	-0.057	0.029	0.050	0.250	0.011	0.042	0.801	1.000

*Note.* Model-based effect sizes are not available for literacy and numeracy skills due to the lack of a measurement model on which to scale and standardize the impact estimates. Instead, this table provides standardized impact estimates as effect sizes.

*Secondary Outcomes.* Twenty-six weeks of access to HCT showed positive impacts on Syrian refugee children's behavioral regulation (adjusted  $ES = 0.31$ ,  $p < .001$ ,  $q = .001$ ). Access to programming did not have any impacts on any other secondary outcomes (see Table 4).

*RQ2: Impact of Skill-Targeted SEL Activities Over 26 weeks (26 weeks of HCT vs. 26 Weeks of HCT+Act)*

*Impacts on Primary Outcomes.* Students with access to 26 weeks of Tutoring in a Healing Classroom plus skill-targeted SEL reported higher school stress compared to those with access to HCT programming without skill-targeted activities (adjusted  $ES = 0.21$ ,  $p = .009$ ,  $q = .099$ ), though this finding is marginally significant after correction for multiple comparisons. There were no other statistically significant impacts on students' perceptions of public school or remedial program environments. We did detect trends that suggest that access to 26 weeks of skill-targeted SEL activities over and above HCT showed a marginal impact on decreasing aggressive reaction (adjusted  $ES = -0.12$ ,  $p = .089$ ,  $q = .669$ ) and higher levels of teacher-reported externalizing behaviors (adjusted  $ES = 0.22$ ,  $p = .009$ ,  $q = .099$ ). However, these results were no longer statistically significant after the FDR correction.

Access to two cycles of HCT programming with skill-targeted SEL activities had no detectable impact on children's overall literacy and numeracy competency nor on discrete literacy and numeracy subtask skills when compared to children with two cycles of HCT programming without skill-targeted SEL activities (see Table 6).

*Sensitivity Analyses.* Table 7 presents adjusted impact estimates from the sensitivity analyses testing the impact of two cycles of skill-targeted SEL activities implemented in HCT classrooms, excluding six dropout sites. We found overall similar patterns of impacts of skill-targeted SEL activities in directions and magnitude—with higher level of school-related stress ( $ES = 0.19$ ,  $p = .023$ ,  $q = .299$ ), decrease in aggressive reaction ( $ES = -0.17$ ,  $p = .028$ ,  $q = .401$ ), and an increase in teacher-reported externalizing behaviors ( $ES = 0.22$ ,  $p = .053$ ,  $q = .401$ )—significant ( $< .05$ ) or marginally significant based on  $p$ -values. However, these results were no longer statistically significant after the FDR correction.

## Discussion

Using a cluster randomized design of non-equivalent treatment groups, this study evaluated the impact of two types of remedial program enhancement for Syrian refugee

TABLE 6

*Impact Effect Size (ES) Estimates of 26 Weeks of Tutoring in a Healing Classroom Plus SEL Activities (2C-HCT+Act) Over and Above 26 weeks of Tutoring in a Healing Classroom Alone (2C-HCT)*

	Unadjusted				Adjusted			
	<i>ES</i>	<i>SE</i>	<i>p</i>	<i>q</i>	<i>ES</i>	<i>SE</i>	<i>p</i>	<i>q</i>
<b>Primary Outcomes</b>								
School-related Stress and Stress Reactivity								
School-related stress	0.223	0.092	0.016	0.289	0.209	0.080	0.009	0.145
Stress reactivity	0.075	0.113	0.504	1.000	0.053	0.091	0.560	1.000
Executive Function								
Working memory	0.009	0.043	0.832	1.000	0.028	0.048	0.557	1.000
Inhibitory control	0.022	0.043	0.607	1.000	0.017	0.042	0.683	1.000
Teacher-Report Social and Emotional Functioning								
Prosocial behavior and social acceptance	0.010	0.148	0.948	1.000	-0.052	0.144	0.719	1.000
Hyperactive and externalizing behaviors	0.206	0.115	0.072	.454	0.222	0.117	0.058	0.606
Emotional distress and internalizing behaviors	0.205	0.210	0.330	1.000	0.224	0.204	0.272	1.000
Working memory functioning	-0.035	0.135	0.793	1.000	-0.096	0.143	0.501	1.000
Inhibitory control functioning	-0.009	0.133	0.948	1.000	-0.052	0.141	0.714	1.000
Cognitive and Emotional Regulation								
Hostile attribution bias	-0.101	0.103	0.326	1.000	-0.090	0.090	0.329	1.000
Anger dysregulation	-0.008	0.114	0.944	1.000	0.033	0.120	0.788	1.000
Sadness dysregulation	0.125	0.106	0.237	1.000	0.163	0.110	0.142	0.711
Aggressive reaction	-0.135	0.073	0.066	0.454	-0.120	0.070	0.089	0.628
Behavioral Regulation	0.014	0.069	0.840	1.000	-0.017	0.060	0.772	1.000
<b>Secondary Outcomes</b>								
Perceptions of Remedial Classroom Environment								
Positive climate	0.065	0.067	0.332	1.000	0.070	0.054	0.194	1.000
Engaging and safe remedial school	-0.006	0.057	0.912	1.000	0.005	0.054	0.992	1.000
Perceptions of Public-School Environment								
Caring and supportive teachers	0.141	0.211	0.505	1.000	0.131	0.187	0.485	1.000
Engaging and motivating school	0.108	0.170	0.525	1.000	0.097	0.139	0.487	1.000
Respectful and inclusive school	0.139	0.241	0.565	1.000	0.137	0.201	0.495	1.000
Safe school	0.037	0.103	0.718	1.000	0.053	0.101	0.599	1.000
Internalizing symptoms (MFQ)	0.078	0.090	0.382	1.000	0.088	0.085	0.297	1.000
Academic Outcomes								
EGRA literacy competence	0.006	0.013	0.627	1.000	0.010	0.011	0.327	1.000
EGMA numeracy competence	-0.039	0.050	0.435	1.000	-0.040	0.049	0.416	1.000

students enrolled in Lebanese public schools: longer program duration (10 vs. 26 weeks) and the addition of skill-targeted SEL activities to 26 weeks of the base remedial tutoring program. To our knowledge, this study is among the first to rigorously evaluate the impact of remedial program design in a crisis and conflict-affected context or with a refugee population. The results suggest that longer duration produced a small increase in students' literacy outcomes, but neither increasing program duration nor the addition of skill-targeted SEL activities had a meaningful impact on students' math outcomes. Furthermore, neither longer duration nor the addition of skill-targeted SEL activities provided strong, consistent evidence of positive impacts on students' well-being, as measured by an array

of social and emotional skills and competencies. The findings suggest that while an academic remedial program of either 10 or 26 weeks that incorporates positive climate principles produced improvement in social and emotional and academic outcomes—with or without the addition of skill-targeted SEL activities—most measurable differences in social and emotional outcomes were generally lower in the program with longer duration. Given the enormity of the refugee population and the limited funding available for education in emergency contexts (Nicolai & Hine, 2015), this study complicates the notion that more program dosage is better and suggests that even short-term remedial programming can generate impacts for vulnerable students.

TABLE 7

*Sensitivity Analyses for the SEL Contrast: Adjusted Impact Estimates Excluding Dropout Clusters at Post-Intervention*

	Targeted SEL Contrast			
	<i>Effect Size</i>	<i>SE</i>	<i>p</i>	<i>q</i>
<b>Primary Outcomes</b>				
School-Related Stress and Stress Reactivity				
School-related Stress	0.194	0.085	0.023	0.244
Stress reactivity	0.008	0.099	0.935	1.000
Cognitive and Emotional Regulation				
Hostile attribution bias	-0.139	0.097	0.152	0.719
Anger dysregulation	-0.021	0.128	0.869	1.000
Sadness dysregulation	0.144	0.129	0.266	0.822
Reactive aggression	-0.170	0.077	0.028	0.244
Behavioral regulation	-0.002	0.061	0.970	1.000
Executive Function				
Working memory	0.050	0.059	0.396	1.000
Inhibitory control	0.023	0.046	0.627	1.000
Teacher-Report Social and Emotional Functioning				
Prosocial behavior and social acceptance	-0.111	0.141	0.431	1
Hyperactive and externalizing behaviors	0.220	0.114	0.053	0.270
Emotional distress and internalizing behaviors	0.224	0.210	0.287	0.822
Working memory functioning	-0.163	0.136	0.233	0.822
Inhibitory control functioning	-0.086	0.149	0.561	1.000
<b>Secondary Outcomes</b>				
Perceptions of Public School Environment				
Caring and supportive teachers	0.202	0.192	0.291	0.536
Engaging and motivating school	0.180	0.137	0.190	0.536
Respectful and inclusive school	0.239	0.213	0.262	0.536
Safe school	0.023	0.113	0.836	0.831
Perceptions of Remedial Classroom Environment				
Positive climate	0.099	0.054	0.066	0.401
Engaging and safe remedial school	0.031	0.056	0.573	0.831
Internalizing Symptoms	0.055	0.094	0.562	0.831
Academic Outcomes				
EGRA literacy competence	0.011	0.011	0.314	0.536
EGMA numeracy competence	-0.060	0.054	0.268	0.536
<b>Supplementary Outcomes</b>				
EGRA Literacy Skills (Subtasks)				
Receptive vocabulary	0.055	0.111	0.621	1.000
Letter recognition	0.121	0.094	0.199	1.000
Grapheme	0.014	0.135	0.920	1.000
Invented word reading	0.028	0.209	0.894	1.000
Oral passage reading	0.292	0.222	0.189	1.000
Reading comprehension	0.024	0.161	0.879	1.000
Word dictation	0.044	0.137	0.751	1.000
EGMA Numeracy Skills (Subtasks)				
Number identification	-0.090	0.129	0.485	0.667
Number discrimination	-0.054	0.126	0.668	0.667
Missing number	-0.147	0.113	0.190	0.667
Addition level 1	-0.148	0.118	0.208	0.667
Addition level 2	-0.139	0.137	0.311	0.667
Subtraction level 1	-0.155	0.133	0.242	0.667
Subtraction level 2	-0.192	0.154	0.214	0.667
Multiplication	-0.185	0.176	0.294	0.667
Word problems	-0.178	0.112	0.113	0.667

Our first research question asked whether a longer duration of remedial programming would significantly impact students' academic skills and perceptions of their remedial and public school climates. Students who had access to remedial programming for 26 weeks (two cycles) showed a small but statistically significant increase in literacy (but not math) skills ( $ES = 0.04$ ) when compared to students with one cycle of access. External observers also reported increased behavioral regulation in students ( $ES = 0.31$ ) after two cycles of access to remedial programming, compared to students with access to one cycle. However, students with two cycles of access reported less positive perceptions of their public school environment as well as their remedial environment than students with access to one cycle. Specifically, students reported less caring and supportive public school teachers ( $ES = -0.89$ ), engaging and motivating public school climate ( $ES = -0.83$ ), and respectful and inclusive public schools ( $ES = -0.87$ ). For remedial sites, students reported lower perceptions of positive climate ( $ES = -0.24$ ) and less engaging and safe remedial class environment ( $ES = -0.15$ ). We found no other differences in academic or social and emotional outcomes for extended program duration.

We note here that the addition of one cycle (16 weeks) of programming has demonstrated positive effects on a performance assessment (EGRA) and a third-party observer (PSRA) but less positive effects on student perceptions of their remedial and public school environments. It may be the case that students are cautious in new environments, resulting in an especially placid—though somewhat artificial—classroom environment for a short period of time, inflating perceptions of classroom climate. It's also possible that waitlisted students and teachers may have been aware of their condition in the first cycle, leading to higher initial enthusiasm and effort once they were granted access to the program, resulting in more positive perceptions. Last, it's possible that student feelings regarding classroom safety and supportiveness may peak short-term, when an environment is new, then decrease as students become accustomed to the setting(s). This hypothesis is supported by prior studies of classroom climate over time (Booth & Gerard, 2014) as well as students' access to a Healing Classrooms approach; Tubbs Dolan et al. (2022) found sizeable positive impacts in increased perceptions of the safety and supportiveness of public schools after one cycle (16 weeks) of Healing Classrooms Tutoring in Syrian Refugees ( $ES = 0.43$ – $0.67$ ), and Torrente et al. (2015) found more positive perceptions of school environment among Congolese students who attended a public school with Healing Classrooms–infused pedagogy ( $ES = 0.22$ ). However, the positive perceptions were not detectable in the Congolese sample after two years of

intervention (Torrente et al., 2019). A better understanding of how student perceptions of classroom climate may shift over time, including whether there is a predictable “pioneering” or “honeymoon” effect of an intervention that targets classroom culture, and how quickly it may temper is an avenue of future research.

Our second research question asked whether the addition of skill-targeted SEL activities over the full 26-week duration (two program cycles) had measurable impacts on targeted social and emotional skills and competencies compared to programming with climate-targeted SEL alone. We found that students with access to two cycles of HCT programming with skill-targeted activities reported higher school-related stress ( $ES = 0.21$ ) compared to students in HCT programming alone, which was marginal ( $q = .09$ ) after FDR correction. While both groups reported decreases in school-related stress from baseline, students in the skill-targeted activities condition declined less than those in the HCT-only condition. This finding is consistent with Tubbs Dolan et al. (2022), who also found a smaller decrease in reports for school-related stress for one cycle of HCT plus skill-targeted activities focused on mindfulness. While mindfulness, in particular, was hypothesized to reduce stress, it's plausible that students became more aware of and more comfortable reporting feelings of stress through participating in activities that explicitly targeted mind-body awareness and emotional regulation. It's also possible that regulatory and stress mechanisms operate differently in conflict-affected contexts marked by chronic daily difficulties (Miller & Rasmussen, 2010).

We found no other statistically significant differences in social and emotional or academic outcomes for students with access to two cycles of skill-targeted SEL activities above and beyond HCT programming. While we hypothesized that increased dosage and modality of SEL programming would increase impact (January et al., 2011), the skills targeted by SEL activities varied by cycle; in cycle one, mindfulness activities targeted stress reduction and mind-body awareness, whereas play-based brain-game activities targeted executive function skills in cycle two. A similar pattern of findings—a positive impact of mindfulness after the first cycle and null effects after the second cycle that tested the cumulative effect of mindfulness and brain games implemented sequentially—is replicated in Niger, where similar HCT programming and SEL activities were implemented (Kim et al., 2023, under review). It's possible that these skill-targeted activities did not reach the threshold of implementation quality or quantity to produce hypothesized effects in the current treatment contrast, over and above the effects of climate-targeted SEL. It's also possible that the impacts generated by different activities targeting different

skills neutralized one another or the activities were not optimally sequenced to produce impact. While activities were sequenced to build skills for physiological down-regulation of emotions prior to addressing more cognitive skills such as working memory and inhibition (Jones et al., 2017), there is little empirical research—even from high-income contexts—to inform whether or how to best sequence SEL skill acquisition (Cipriano et al., 2023; Lawson et al., 2019). Last, it's also possible that the treatment contrast was not sufficiently large—particularly given the climate-targeted SEL provided in the HCT condition, which is designed to provide safe, supportive, and predictable learning environments, or that implementation factors such as low attendance and short duration of the second cycle attenuated any possible impacts.

### *Limitations*

We note several limitations of the current study. First, we evaluated a specific version of remedial tutoring—the International Rescue Committee's Healing Classrooms tutoring—in this study, which may differ in important ways from other remedial tutoring programs. For example, the base program (HCT) infuses academic instruction with climate-targeted social and emotional practices, and teachers are given ongoing professional development, which may not be provided in other remedial tutoring programs. Furthermore, with the exception of program duration and the addition of skill-targeted SEL skills, we cannot be sure which programmatic component—or combination thereof—produced the results of this study. We therefore caution that these results are not generalizable beyond HCT remedial tutoring in Lebanon. Second, recruitment for the program was limited to Syrian refugee children enrolled in Lebanese public schools in accordance with Lebanon MEHE policies. Therefore, the findings are not generalizable to the approximately 250,000 out-of-school Syrian refugee children in Lebanon. Future studies could explore whether the findings hold both for these populations within Lebanon as well as in other contexts with different demographics and cultural backgrounds. Third, this paper focused on average impacts and did not explore variation of impacts on different subgroups with varying level of vulnerability. Such subgroup analyses hold substantial importance, both in a general and perhaps more so in research within resource-constrained environments, and may reveal patterns that can explain the unexpected findings presented in this paper. However, given the complexity of the analysis presented here, to do justice such additional analysis deserves substantial additional analysis and a careful interpretation of results, which is beyond the scope of this paper. We hope to explore this fruitful avenue in our future research.

In addition, the comparison of three active treatment arms has advantages and limitations compared to other study designs. The most obvious limitation is that we cannot know

the full impact of the intervention in the absence of treatment without a no-treatment control group. However, there are ethical issues to consider with a no-treatment control group, particularly with interventions related to desperately needed educational and well-being support (Saks et al., 2002). Moreover, the budgetary and programmatic realities on the ground, in addition to the ethical imperative, necessitated providing all participating students with access to remedial tutoring. Nonetheless, the study design resulted in several threats to internal validity.

First, we note that our test of extending programming from 10 to 26 weeks has several confounding factors. Students who had access to the full 26 weeks (two cycles) of programming had access from November to June, while those with access for 10 weeks had access from March to June. To the degree that seasonal variation in teaching and learning in either nonformal remedial or formal public school environments occurs (Plank & Condliffe, 2013; Scales et al., 2020), this variation could be captured as part of our impact estimates. Second, students who progressed from the first to second cycle experienced a progression of curriculum that increased in difficulty. While this is a feature of most academic curricula, it's worth noting, given the findings, that a shorter duration of programming produced more favorable student perceptions of classroom climate. Third, six sites randomized to receive access to two cycles of HCT, plus skill-targeted SEL dropped out for reasons described previously, and one was lost to follow-up due to insecure safety conditions. Though these sites are included in the data collection and analysis to ensure the most rigorous impact estimates, the lack of treatment as well as the shortened assessment time (compared to program time) due to Ramadan results in a downward bias of our impact estimates and may inform nonsignificant findings.

### *Implications and Recommendations*

Despite these limitations, this study makes several important contributions. First, its experimental design allows for causal inference that is rare in educational programming in conflict and crisis-affected contexts like Lebanon. We know of few studies that rigorously evaluate remedial educational programming in such contexts—despite its widespread implementation—and know of none that have specifically investigated design features of the programming. Second, this study found that educational support services—such as remedial tutoring with SEL—as brief in duration as 10 weeks can improve students' perceptions of their public school, improve academic outcomes, and increase student well-being. In a public school environment marked by high rates of violence and bullying, large class sizes, and low levels of instruction, such a boost could translate to a meaningful impact on refugee students (Abdul-Hamid & Yassine, 2020). Third, the evidence generated from this design is a

first step toward providing policymakers and practitioners with necessary information regarding how to deploy scarce educational programming resources in areas of conflict and crisis for vulnerable refugee students. While we hypothesized that increasing the duration of programming and enhancing it with skill-targeted SEL activities would increase impacts, the results are more complicated; instead, they suggest that in some cases, resources may be best deployed by increasing the number of students granted access to short-term programming rather than increasing the duration of said programming. However, we caution against over-interpreting these study results, as short-term assessments to evaluate program impact may be misleading, and repeated evaluations of such programming over longer time periods will likely provide additional and potentially different information (King & Behrman, 2009). Replication of studies evaluating the impact of program duration, inclusive of its cost-effectiveness, are fruitful avenues for future research and would support policymakers and NGOs to make informed decisions about how to allocate limited resources for maximum impact.

Last, this study measures the impact of access to remedial tutoring programming on a robust set of academic, social, emotional, and cognitive skills and outcomes. Measures used a variety of reporting formats (e.g., direct assessment, observation, self, and other-report survey) and reporters (e.g., children, parents, and enumerators), and program impacts were estimated using rigorous modeling approaches that account for missing data and measurement invariance. Given growing interest in measuring and improving complex holistic learning outcomes, such knowledge of and attention to measurement and psychometric best practices is critical.

Given this evidence, we recommend that governments and NGOs working with refugee populations prioritize the provision of remedial educational programming and consider their context's tradeoffs of implementing social-emotional learning activities within the remedial education programming. We also recommend that governments and NGOs consider the duration of remedial education programming. The study suggests that while longer programs may be more effective in some contexts, in others it may be more beneficial to increase the number of students granted access to short-term programming. Therefore, program implementers should conduct context-specific needs assessments to determine the most effective duration of remedial education programming for their refugee populations. Lastly, given that climate- and skill-targeted SEL-infused remedial programs can impact a range of academic, social, and emotional outcomes, governments and NGOs should consider adopting a holistic measurement approach that recognizes the interconnectedness of these domains and seeks to improve all aspects of children's learning and well-being.

## Conclusion

This study investigates two policy-relevant, field-feasible enhancements of remedial programming: duration and the addition of skill-targeted SEL activities. Broadly, these enhancements were found to improve children's basic functioning in literacy and behavioral regulation, as measured by performance assessments and external observations of behavior but *adversely* affect children's perceptions of their learning environments and self-reported academic stress. This study complicates the notion that "more is better" when it comes to dosage of academic remedial programming—potentially leading to increased effectiveness of programmatic funds for program implements and governments—while providing sufficient evidence to continue the implementation and investigation of remedial programming and social-emotional learning with refugee populations.

## Acknowledgments

We wish to thank the International Rescue Committee, especially its Education and Research, Evaluation and Learning Units, and its education, research, and M&E staff in Lebanon for their commitment to learning how best to support children in crisis contexts. Finally, we wish to thank the Syrian children, parents, and teachers in Lebanon for allowing us to engage them in this work.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We also thank Mayari Montes de Oca and Patrick Anker at New York University's Global TIES for Children for data support; Dubai Cares, the Spencer Foundation, and an anonymous donor for direct financial support that enabled the larger study of which this paper is a part; and the NYU Abu Dhabi Research Institute for supporting the time of Drs. Kim, Gjicali, and Aber to work on this paper.

## Open Practices

The data for this article can be found at doi: 10.7910/DVN/97Q2B8.

## ORCID iDs

Lindsay Brown  <https://orcid.org/0000-0003-0934-1736>

Ha Yeon Kim  <https://orcid.org/0000-0001-7468-1310>

Carly Tubbs Dolan  <https://orcid.org/0000-0002-8589-8248>

## Notes

1. An additional 11 sites were included as waitlist control in the Cycle 1 sample. These 11 sites were then randomized to implement HCT+brain games for a separate pilot study, which is reported elsewhere.

2. This decision was made by program implementers at the field level. The initial program evaluation was focused on Cycle 1. In order to provide programming to the additional 21 sites that were waitlisted in Cycle 1, our partners intended to discontinue programming at 20 of 33 2C-HCT+Act sites, as the additional SEL activities were perceived by the programming staff as receiving more community support compared to HCT-only sites. After consultation, we were able to maintain 14 sites out of 20 sites for a total loss of 6 sites.

3. Anecdotal field reports suggested that teachers were incentivized to enroll a minimum number of students in order to ensure their job security, which may have resulted in the registration of students who did not plan to attend the program.

4. The impact estimates of 2C-HCT+Act are over and above the changes found in the 2C-HCT-only group—that is, the changes in 2C-HCT+Act group from baseline to endline is 2C-HCT group change + impact estimate of 2C-HCT+Act group.

## References

- Abdul-Hamid, H., & Yassine, M. (2020). *Political economy of education in Lebanon: Research for results program*. World Bank. <https://doi.org/10.1596/978-1-4648-1546-1>
- Aber, J. L., Torrente, C., Starkey, L., Johnston, B., Seidman, E., Halpin, P., Shivshanker, A., Weisenhorn, N., Annan, J., & 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*(3), 507–529. <https://doi.org/10.1080/19345747.2016.1236160>
- Adelman, E. (2018). *Challenges of integration, obligation, and identity: Exploring the experiences of teachers working to educate Syrian refugee children in Lebanon* [Doctoral Dissertation, Harvard University].
- Agostinelli, F., Avitabile, C., Bobba, M., & Sanchez, A. (2019). *The short-term effects of the mobile pedagogical tutors: Evidence from a randomized control trial in rural Mexico*. Mimeo.
- Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, 1–10.
- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India\*. *The Quarterly Journal of Economics, 122*(3), 1235–1264. <https://doi.org/10.1162/qjec.122.3.1235>
- Banerji, R., Bhattacharjee, S., & Wadhwa, W. (2013). The annual status of education report (ASER). *Research in Comparative and International Education, 8*(3), 387–396. <https://doi.org/10.2304/rcie.2013.8.3.387>
- Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*(3), 491–507. <https://doi.org/10.1093/biomet/93.3.491>
- Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: Experimental canalization of brain and behavior. *American Psychologist, 67*(4), 309–318. <https://doi.org/10.1037/a0027493>
- Booth, M. Z., & Gerard, J. M. (2014). Adolescents’ stage-environment fit in middle and high school: The relationship between students’ perceptions of their schools and themselves. *Youth & Society, 46*(6), 735–755. <https://doi.org/10.1177/0044118X12451276>
- Brown, L. E., Kim, H. Y., Tubbs Dolan, C., Brown, A., Sklar, J., & Aber, J. L. (2023). Remedial programming and skill-targeted SEL in low-income and crisis-affected contexts: Experimental evidence from Niger. *Journal of Research on Educational Effectiveness, 16*(4), 583–614. <https://doi.org/10.1080/19345747.2022.2139785>
- Brudevold-Newman, A., de Hoop, T., Holla, C., Isaboke, D., Kinoti, T., Ring, H., & Rothbard, V. (2023). The effects of booster classes in protracted crisis settings: Evidence from Kenyan refugee camps. *Journal of Development Effectiveness, 15*(3), 287–301. <https://doi.org/10.1080/19439342.2023.2229294>
- Buckner, E., Spencer, D., & Cha, J. (2018). Between policy and practice: The education of Syrian refugees in Lebanon. *Journal of Refugee Studies, 31*(4), 444–465. <https://doi.org/10.1093/jrs/fex027>
- Chiesa, A., Serretti, A., & Jakobsen, J. C. (2013). Mindfulness: Top-down or bottom-up emotion regulation strategy? *Clinical Psychology Review, 33*(1), 82–96. <https://doi.org/10.1016/j.cpr.2012.10.006>
- Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development, 94*(5), 1181–1204. <https://doi.org/10.1111/cdev.13968>
- Connor-Smith, J. K., Compas, B. E., Wadsworth, M. E., Thomsen, A. H., & Saltzman, H. (2000). Responses to stress in adolescence: Measurement of coping and involuntary stress responses. *Journal of Consulting and Clinical Psychology, 68*(6), 976–992.
- Di Giunta, L., Iselin, A.-M. R., Eisenberg, N., Pastorelli, C., Gerbino, M., Lansford, J. E., Dodge, K. A., Caprara, G. V., Bacchini, D., Uribe Tirado, L. M., & Thartori, E. (2017). Measurement invariance and convergent validity of anger and sadness self-regulation among youth from six cultural groups. *Assessment, 24*(4), 484–502. <https://doi.org/10.1177/1073191115615214>
- Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., Tirado, L. M. U., Zelli, A., Alampay, L. P., Al-Hassan, S. M., Bacchini, D., Bombi, A. S., Bornstein, M. H., Chang, L., Deater-Deckard, K., Di Giunta, L., Oburu, P., & Pastorelli, C. (2015). Hostile attributional bias and aggressive behavior in global context. *Proceedings of the National Academy of Sciences, 112*(30), 9310–9315. <https://doi.org/10.1073/pnas.1418572112>
- Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development, 40*, 315–322. <https://doi.org/10.1016/j.ijedudev.2014.11.004>
- 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: Social and emotional learning. *Child Development, 82*(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45*(3–4), 294–309. <https://doi.org/10.1007/s10464-010-9300-6>

- Finch, J. E., Wolf, S., & Lichand, G. (2022). Executive functions, motivation, and children's academic development in Côte d'Ivoire. *Developmental Psychology*, *58*(12), 2287–2301. <https://doi.org/10.1037/dev0001423>
- 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*(1), 141–157. <https://doi.org/10.1177/1745499919829217>
- Gjicali, K., Wu, Z., & Kim, H. Y. (2020a). *Education in emergencies: Evidence for Action (3EA) Lebanon 2016-17: Psychometric report of measures for academic, cognitive, social and emotional skills*. [Technical Report]. Global TIES for Children, New York University. <https://doi.org/10.6084/m9.figshare.24107898>
- Gjicali, K., Wu, Z., Kim, H. Y., & Tubbs Dolan, C. (2020b). *Psychometric evidence on the Child Friendly School Questionnaire for Syrian Children in Lebanon (CFSQ-SL): A measurement tool of student-perceived school climate* [Technical working paper]. NYU Global TIES for Children. <https://inee.org/resources/child-friendly-school-questionnaire-syrian-children-lebanon-cfsq-sl>
- Glewwe, P. (2013). *Education policy in developing countries*. University of Chicago Press.
- Global Education Monitoring Report Team (GEMR). (2019). *Education as healing: Addressing the trauma of displacement through social and emotional learning—UNESCO Digital Library* (Policy Paper, 38). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000367812>
- Godfrey, E. B., Osher, D., Williams, L. D., Wolf, S., Berg, J. K., Torrente, C., Spier, E., & Aber, J. L. (2012). Cross-national measurement of school learning environments: Creating indicators for evaluating UNICEF's Child Friendly Schools Initiative. *Children and Youth Services Review*, *34*(3), 546–557. <https://doi.org/10.1016/j.childyouth.2011.10.015>
- Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13473–13480. <https://doi.org/10.1073/pnas.93.24.13473>
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, *177*(6), 534–539. <https://doi.org/10.1192/bjp.177.6.534>
- Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy* (1st ed.). RTI Press. <https://doi.org/10.3768/rtipress.2011.bk.0007.1109>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education*, *84*(1), 175–196. <https://doi.org/10.1080/00220973.2014.952397>
- Human Rights Watch. (2016). *Growing up without an education*. <https://www.hrw.org/report/2016/07/19/growing-without-education/barriers-education-syrian-refugee-children-lebanon>
- January, A. M., Casey, R. J., & Paulson, D. (2011). A meta-analysis of classroom-wide interventions to build social skills: Do they work? *School Psychology Review*, *40*(2), 242–256. <https://doi.org/10.1080/02796015.2011.12087715>
- Jones, D. E., Greenberg, M., & Crowley, M. (2015a). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, *105*(11), 2283–2290. <https://doi.org/10.2105/AJPH.2015.302630>
- Jones, S. M., Bailey, R., & Barnes, S. (2015b). *Classroom Executive Function Survey (CEFS): A measure of executive function and self-regulation skills in everyday classroom behavior*. Harvard University.
- Jones, S. M., Barnes, S. P., Bailey, R., & Doolittle, E. J. (2017). Promoting social and emotional competencies in elementary school. *The Future of Children*, *27*(1), 49–72.
- Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies and commentaries. *Social Policy Report*, *26*(4), 1–33. <https://doi.org/10.1002/j.2379-3988.2012.tb00073.x>
- Keresteš, G. (2006). Children's aggressive and prosocial behavior in relation to war exposure: Testing the role of perceived parenting and child's gender. *International Journal of Behavioral Development*, *30*(3), 227–239. <https://doi.org/10.1177/0165025406066756>
- 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. <https://doi.org/10.1016/j.appdev.2020.101142>
- Kim, H. Y., Brown, L. E., Tubbs Dolan, C., Ngoga, E., Fernandes, R. F., & Aber, J. L. (under review, 2023). School-based universal social and emotional learning interventions for education in conflict and crisis: Experimental impacts of in Niger.
- Kim, H. Y., Wu, Z., Gjicali, K., & Tubbs Dolan, C. (2021). *Psychometric evidence on a measure of school-related stress and stress reactivity among Syrian children in Lebanese public schools: Response to Stress Questionnaire (RSQ)* [Technical working paper]. NYU Global TIES for Children. <https://inee.org/resources/response-stress-questionnaire-rsq>
- King, E. M., & Behrman, J. R. (2009). Timing and duration of exposure in evaluations of social programs. *The World Bank Research Observer*, *24*(1), 55–82. <https://doi.org/10.1093/wbro/lkn009>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Press.
- Lawson, G. M., McKenzie, M. E., Becker, K. D., Selby, L., & Hoover, S. A. (2019). The core components of evidence-based social emotional learning programs. *Prevention Science*, *20*(4), 457–467. <https://doi.org/10.1007/s11121-018-0953-y>
- 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*(1), 227–257. <https://doi.org/10.1146/annurev-psych-120710-100356>
- McCoy, D. C., Zuilkowski, S. S., Yoshikawa, H., & Fink, G. (2017). Early childhood care and education and school readiness in Zambia. *Journal of Research on Educational Effectiveness*, *10*(3), 482–506. <https://doi.org/10.1080/19345747.2016.1250850>
- Mendenhall, M. A., Russell, S. G., & Bruckner, E. (2017). *Urban refugee education: Strengthening policies and practices for access, quality, and inclusion*. <https://doi.org/10.7916/D8R49WGP>

- Miller, K. E., & Rasmussen, A. (2010). War exposure, daily stressors, and mental health in conflict and post-conflict settings: Bridging the divide between trauma-focused and psychosocial frameworks. *Social Science & Medicine*, *70*(1), 7–16. <https://doi.org/10.1016/j.socscimed.2009.09.029>
- Ministry of Education and Higher Education. (2018). *RACE II fact sheet: November 2018*. <http://racepmulebanon.com/images/fact-sheet-november-2018.pdf>
- Muthén, L. K., & Muthén, B. (2018). Mplus. *The Comprehensive Modelling Program for applied researchers: User's guide*, 5. Muthén & Muthén.
- Nicolai, S., & Hine, S. (2015). *Investment for education in emergencies: A review of evidence*. ODI.
- OECD. (2017). *Setting the stage: Proceedings of the first policy forum (Strength Through Diversity)*. <http://www.oecd.org/education/school/Proceedings-May2017.pdf>
- Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal*, *50*(5), 1152–1182.
- Reed, R. V., Fazel, M., Jones, L., Panter-Brick, C., & Stein, A. (2012). Mental health of displaced and refugee children resettled in low-income and middle-income countries: Risk and protective factors. *The Lancet*, *379*(9812), 250–265. [https://doi.org/10.1016/S0140-6736\(11\)60050-0](https://doi.org/10.1016/S0140-6736(11)60050-0)
- Ruauadel, H., & Morrison-Métois, S. (2017). *Responding to refugee crises in developing countries: What can we learn from evaluations?* <https://doi.org/10.1787/ae4362bd-en>
- Saavedra, J. E., Näslund-Hadley, E., & Alfonso, M. (2019). Remedial inquiry-based science education: Experimental evidence from Peru. *Educational Evaluation and Policy Analysis*, *41*(4), 483–509. <https://doi.org/10.3102/0162373719867081>
- Saks, E. R., Jeste, D. V., Granholm, E., Palmer, B. W., & Schneiderman, L. (2002). Ethical issues in psychosocial interventions research involving controls. *Ethics & Behavior*, *12*(1), 87–101. [https://doi.org/10.1207/S15327019EB1201\\_6](https://doi.org/10.1207/S15327019EB1201_6)
- Scales, P. C., Pekel, K., Sethi, J., Chamberlain, R., & Van Boekel, M. (2020). Academic year changes in student-teacher developmental relationships and their linkage to middle and high school students' motivation: A mixed methods study. *The Journal of Early Adolescence*, *40*(4), 499–536. <https://doi.org/10.1177/0272431619858414>
- Shonkoff, J. P., Garner, A. S., Dobbins, M. I., Earls, M. F., Garner, A. S., McGuinn, L., Pascoe, J., & Wood, D. L. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, *129*(1), e232–e246. <https://doi.org/10.1542/peds.2011-2663>
- Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academies Press (US).
- Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. <https://doi.org/10.1037/h0020586>
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173–187. <https://doi.org/10.1016/j.ecresq.2007.01.002>
- Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The Future of Children*, *26*, 57–74.
- Spybrook, J. K., Raudenbush, W., & Liu, X. (2006). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Survey Research Center of the Institute of Social Research at University of Michigan.
- Steuerle, C. E., Reynolds, G., & Carasso, A. (2007). *Investing in Children*, Urban Institute.
- Suntheimer, N. M., Wolf, S., Sulik, M. J., Avornyo, E. A., & Obradović, J. (2022). Executive function mediates the association between cumulative risk and learning in Ghanaian schoolchildren. *Developmental Psychology*, *58*, 1500–1511. <https://doi.org/10.1037/dev0001372>
- Tavitian, L., Atwi, M., Bawab, S., Hariz, N., Zeinoun, P., Khani, M., & Maalouf, F. T. (2014). The Arabic Mood and Feelings Questionnaire: Psychometrics and validity in a clinical sample. *Child Psychiatry & Human Development*, *45*(3), 361–368. <https://doi.org/10.1007/s10578-013-0406-6>
- Torrente, C., Aber, J. L., Starkey, L., Johnston, B., Shivshanker, A., Weisenhorn, N., Annan, J., Seidman, E., Wolf, S., & Tubbs Dolan, C. (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*(3), 413–447. <https://doi.org/10.1080/19345747.2018.1561963>
- Torrente, C., Johnston, B., Starkey, L., Seidman, E., Shivshanker, A., Weisenhorn, N., Annan, J., & Aber, J. L. (2015). Improving the quality of school interactions and student well-being: Impacts of one year of a school-based program in the Democratic Republic of Congo. *The Journal on Education in Emergencies*, *1*(1), 48–91. <https://doi.org/10.17609/N8NH2N>
- Tubbs Dolan, C., Kim, H. Y., Brown, L., Gjicali, K., Borsani, S., Houchaimi, S. E., & Aber, J. L. (2022). Supporting Syrian Refugee children's academic and social-emotional learning in national education systems: A cluster randomized controlled trial of nonformal remedial support and mindfulness programs in Lebanon. *American Educational Research Journal*, *59*(3), 419–460.
- UNESCO. (2017). More than one-half of children and adolescents are not learning worldwide. *UIS Fact Sheet No. 46*. <https://uis.unesco.org/sites/default/files/documents/fs46-more-than-half-children-not-learning-en-2017.pdf>
- UNESCO, UNICEF, & World Bank. (2020). *What have we learnt? Overview of findings from a survey of ministries of education on national responses to COVID-19*. [https://uis.unesco.org/sites/default/files/documents/national-education-responses-to-covid-19-web-final\\_en\\_0.pdf](https://uis.unesco.org/sites/default/files/documents/national-education-responses-to-covid-19-web-final_en_0.pdf)
- UNHCR. (2017). *UNHCR seeing significant returns of internally displaced amid Syria's continuing conflict*. <https://www.unhcr.org/news/briefing/2017/6/595612454/unhcr-seeing-significant-returns-internally-displaced-amid-syrias-continuing.html>
- Wen, X., Leow, C., Hahs-Vaughn, D. L., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's school performance in kindergarten. *Early Childhood Research Quarterly*, *27*(4), 684–694. <https://doi.org/10.1016/j.ecresq.2011.07.006>
- Wu, Z., Gjicali, K., Kim, H. Y., & Tubbs Dolan, C. (2020). *Self-Regulation Assessment-Assessor Report (SRA-AR)*:

*Psychometric evidence from Syrian refugee children in Lebanon* [Technical working paper]. NYU Global TIES for Children. <https://inee.org/resources/self-regulation-assessment-assessor-report-sra-ar>

### **Authors**

LINDSAY BROWN is a senior research scientist at New York University's Global TIES for Children; email: [lindsay.brown@nyu.edu](mailto:lindsay.brown@nyu.edu). Her work focuses on the design, implementation, and evaluation of flexible and scalable teaching and coaching strategies for global contexts. She is a former middle school teacher.

KALINA GJICALI is a quantitative user-experience researcher at Meta working with cross-functional product teams to improve community messaging experiences on Instagram and Facebook for billions of people around the world.

HA YEON KIM is a senior research scientist at NYU Global TIES for Children in New York; email: [hayeon@nyu.edu](mailto:hayeon@nyu.edu). She is a developmental and educational psychologist interested in measuring, examining, and supporting academic and social-emotional learning of children in low-resource settings and crisis-affected contexts. Her current work includes research on effective ways to support children's learning and wellbeing in conflict-affected contexts across the Middle East, Sub-Saharan Africa, and South and Southeast Asia.

CARLY TUBBS DOLAN is a deputy director at New York University's Global TIES for Children, an international research

center she helped found and launch in New York; email: [carly.tubbs@nyu.edu](mailto:carly.tubbs@nyu.edu). She leads a portfolio of research focused on education measurement, systems, and scaling in contexts of displacement.

PAUL ST JOHN FRISOLI is a senior program specialist at the LEGO Foundation, Denmark; email: [paul.frisoli@legofoundation.com](mailto:paul.frisoli@legofoundation.com). His programmatic and research experiences focus on education in conflict in crisis, mental health and psychosocial support, social-emotional learning, and teacher professional development. His current work focuses on culturally relevant and appropriate play-based approaches to children's development and social-emotional outcomes.

MAHMOUD BWARY is a program officer and education field sector coordinator at UNICEF North and Akkar field office; email: [mbwary@unicef.org](mailto:mbwary@unicef.org). His programmatic and research experience focuses on education for marginalized populations within Lebanon and the MENA region, including supporting in the study design, site recruitment, and programmatic quality of the study in this paper.

J. LAWRENCE ABER is Willner Family Professor in Psychology and Public Policy at NYU Steinhardt and university professor, New York University; email: [la39@nyu.edu](mailto:la39@nyu.edu). He co-directs the international research center Global TIES for Children where he also conducts basic and applied research on the learning and development of children in low-income and conflict-affected countries.