



Developing and Validating the International Social and Emotional Learning Assessment:  
Evidence from a Pilot Test with Syrian Refugee Children in Iraq

Author(s): Nikhit D'Sa and Allyson Krupar

Source: *Journal on Education in Emergencies*, Vol. 7, No. 2 (December 2021), pp. 20-56

Published by: Inter-agency Network for Education in Emergencies

Stable URL: <http://hdl.handle.net/2451/63537>

DOI: <https://doi.org/10.33682/xdpq-bwp2>

## REFERENCES:

This is an open-source publication. Distribution is free of charge. All credit must be given to authors as follows:

D'Sa, Nikhit, and Allyson Krupar. 2021. "Developing and Validating the International Social and Emotional Learning Assessment: Evidence from a Pilot Test with Syrian Refugee Children in Iraq." *Journal on Education in Emergencies* 7 (2): 20-56. <https://doi.org/10.33682/xdpq-bwp2>.

---

The *Journal on Education in Emergencies (JEiE)* publishes groundbreaking and outstanding scholarly and practitioner work on education in emergencies (EiE), defined broadly as quality learning opportunities for all ages in situations of crisis, including early childhood development, primary, secondary, non-formal, technical, vocation, higher and adult education.

Copyright © 2021, Inter-agency Network for Education in Emergencies.



The *Journal on Education in Emergencies*, published by the [Inter-agency Network for Education in Emergencies \(INEE\)](#), is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#), except where otherwise noted.

# DEVELOPING AND VALIDATING THE INTERNATIONAL SOCIAL AND EMOTIONAL LEARNING ASSESSMENT: EVIDENCE FROM A PILOT TEST WITH SYRIAN REFUGEE CHILDREN IN IRAQ

NIKHIT D'SA AND ALLYSON KRUPAR

## ABSTRACT

*The growing focus on social and emotional learning for children of primary grade age in conflict-affected and fragile contexts necessitates an understanding of the effects these programs have. However, the dearth of valid and reliable measures of social and emotional learning skills in low-resource and crisis contexts has constrained the generation of this evidence. The few tools that have robust psychometric properties were developed for use in high-resource contexts; they often have usage costs, limit adaptation, and focus on adults as respondents. To address this gap, we developed the International Social and Emotional Learning Assessment (ISELA), an adaptable, cost-free, open-source, performance-based measure of self-concept, stress management, perseverance, empathy, and conflict resolution in children between ages 6 and 12. In this study, we focused on establishing the validity and reliability of the ISELA when used with Syrian refugee children in Iraq. We tested the latent structure, criterion validity, internal consistency reliability, and interrater reliability of the ISELA with 620 Syrian children. We were able to establish a theoretically grounded factor structure for all of the skills except perseverance. The ISELA can be used reliably by groups of assessors (Krippendorf's  $\alpha > .86$ ) with strong internal consistency (KR-20  $> .70$ ). Our findings for criterion validity were promising but preliminary; grade and exposure to interpersonal threats demonstrated a positive association with social and emotional learning skills.*

Received December 26, 2019; revised February 2, 2021; accepted August 2, 2021; electronically published December 2021.

*Journal on Education in Emergencies*, Vol. 7, No. 2

Copyright © 2021 by the Inter-agency Network for Education in Emergencies (INEE).

ISSN 2518-6833

## INTRODUCTION

Forced displacement due to armed conflict can have deleterious effects on children's psychological and social development, often resulting in issues such as anxiety, depression, interpersonal violence, and posttraumatic stress disorder (Morgos, Worden, and Gupta 2007; Smith et al. 2002; Upadhyay, Srivastava, and Paul 2017). Even children who do not directly experience violence and loss due to forced displacement can be negatively affected by interpersonal aggression, lack of resources, and overcrowded living arrangements (Miller and Rasmussen 2010). Recent developments in neuropsychology reveal that children exposed to severe adversity often experience a physiological stress response that alters how their brains process information, which can adversely affect their ability to learn and thrive (Anda et al. 2011). However, there is preliminary and promising evidence that children who experience adversity can display remarkable psychosocial resilience and academic achievement when provided with opportunities for social and emotional learning (SEL), which is children's ability to understand and manage emotions, feel and show empathy, and develop positive relationships (CASEL 2015; Ungar et al. 2017; Winthrop and Kirk 2008). This has resulted in a proliferation of programs (McNatt et al. 2018) in conflict-affected and fragile states (CAFS) that focus on SEL. However, limited knowledge is available on the impact of these programs and only a few valid and reliable tools measure the SEL skills of children in CAFS. To address this gap, we developed the International Social and Emotional Learning Assessment (ISELA), an adaptable, cost-free, open-source, performance-based measure of self-concept, stress management, perseverance, empathy, and conflict resolution in children between ages 6 and 12. In this paper, we discuss the ISELA's psychometric properties when used with Syrian Kurdish refugee children living in Iraq.

## SEL IN THE CONTEXT OF FORCED DISPLACEMENT

SEL programs have had a demonstrable positive impact on children's academic achievement (Durlak et al. 2011), and they also have been linked to reduced student attrition (Wang et al. 2016), conduct problems, and emotional distress in school-age children (Payton et al. 2008; Taylor et al. 2017). More notably, SEL programs are especially effective in protecting children who have experienced severe adversity and have limited access to other resources (Greenberg et al. 2017; Jones, Greenberg, and Crowley 2015). However, most SEL program impact studies have been conducted in high-resource contexts. In previous reviews and meta-analyses of SEL interventions (Durlak et al. 2011; Puerta, Valerio, and Bernal

2016; Sklad et al. 2012), only 7 percent to 12 percent of the studies were conducted outside of North America and Europe, and none focused on CAFS. This gap in our knowledge about the effects of SEL and other education programs has led the education in emergencies sector to call for more robust and rigorous research that is focused on identifying the impact these programs have on children's short- and long-term outcomes (Burde et al. 2017).

Rigorous efforts have been made recently to study the impact of SEL programs in CAFS. These studies have demonstrated three main challenges. First, they have found that the programs have a limited impact on children's SEL skills. For example, a rigorous cluster-randomized trial of the Learning to Read in a Healing Classroom program in the Democratic Republic of the Congo found that its statistically or practically meaningful impact on children's SEL skills was limited (Aber et al. 2017; Torrente et al. 2019). Even when programs have had a demonstrated impact on children's psychosocial development, the effects have been limited. An evaluation of the Writing for Recovery program in Gaza found that it had a small effect on depression symptoms but no other mental health outcomes (Lange-Nielsen et al. 2012). Second, most of the studies that found a demonstrated impact have been exploratory. For example, the evaluation of a psychoeducation program in Gaza was focused on only sixteen classrooms in four schools (Qouta et al. 2012).

Third, SEL impact studies have focused primarily on distal outcomes related to psychopathology by measuring anxiety, depression symptoms, posttraumatic stress symptoms, and psychological distress. The theory of change of most SEL programs is that these distal psychopathological outcomes are affected by children's more proximal SEL skills, such as self-awareness and perspective-taking (Torrente et al. 2019). However, because of the availability of well-validated tools that focus on children's mental health, research has focused largely on mental health outcomes and used them as a proxy for children's SEL skills. One measure that is often used in studies of children's SEL in CAFS is the Strengths and Difficulties Questionnaire (SDQ), an assessment that has been validated in several CAFS (Woerner et al. 2004). The SDQ was designed primarily as an assessment of children's mental health (Goodman and Goodman 2009) and often is used to screen children into programs of varying intensity (Goodman 1997). However, its use as a measure of children's SEL means that the research is not focused on the children's actual SEL skills; if a tool was not designed to fulfill its intended purpose, programs can make erroneous decisions about its effectiveness (D'Sa 2019b).

### CHALLENGES IN MEASURING SEL

One challenge of responding to the need for further evidence on the effects of SEL programs in CAFS is the dearth of valid, reliable, and feasible measures. If researchers want to focus on measuring the SEL skills of children in CAFS, they face the reality that a majority of the SEL tools were developed in high-resource contexts. A recent review (Halle and Darling-Churchill 2016) noted that, of the 75 SEL measures included, most were developed in North America and only one-third had been used in a language other than English. Child development is mediated by local cultural and social norms (Super and Harkness 1986); thus, it differs according to the setting. Henrich and colleagues (Henrich, Heine, and Norenzayan 2010) gathered findings from across a range of disciplines to demonstrate how samples predominantly drawn from Western, educated, industrialized, rich, and democratic societies are used erroneously to make universal claims. This includes claims about the psychosocial skills targeted by SEL programs. By assuming that attributes of healthy child development look the same across diverse contexts and measuring them within these Western paradigms, we could be incorrectly pathologizing locally normative aspects of development. Or, conversely, we may be failing to identify signs of distress among children in CAFS.

An additional concern is that only a handful of SEL tools developed in high-resource contexts are valid—that is, they measure the intended construct—and reliable—they measure the same construct with consistency. Halle and Darling-Churchill (2016) found that only 6 of 75 measures they reviewed met their validity and reliability criteria. While burgeoning efforts are being made to develop a library of psychometrically rigorous SEL tools that can be used in CAFS (Ferrás, Weiss-Yagoda, and Dolan 2019), there is still a dearth of well-validated measures. This leaves researchers who are studying SEL programs in CAFS with two options. First, adapt and iteratively test the validity and reliability of a measure that was developed in a high-resource context, and attempt to align the constructs and skills with the program being evaluated. Second, develop a custom measure from scratch. Both options are challenging because they require a considerable investment of time and resources (D'Sa 2019b). Moreover, researchers who decide to adapt and iteratively test the validity and reliability of a measure that was developed in a high-resource context must also consider the fact that SEL tools that demonstrate strong validity and reliability (Halle and Darling-Churchill 2016) often require a subscription or must be purchased, and most have restrictive copyrights that do not allow users to freely contextualize the assessment.

Another important limitation is that few validated tools collect SEL information directly from children. The respondents for most well-validated measures of SEL skills for children in primary school grades are caregivers or teachers. In previous reviews of SEL measures (Denham, Ji, and Hamre 2010; Halle and Darling-Churchill 2016), all the measures identified as valid and reliable used caregivers' or teachers' responses to assess children's skills. While caregivers and teachers can provide important information about children's SEL skills, they often offer differing information. Caregivers do not have the opportunity to compare the child in their care to a large sample of children, and teachers, who often have a limited perspective on children in nonacademic settings, do not always know children well at the start of a school year when baseline or pretest data are collected (Darling-Churchill and Lippman 2016). The difficulty in obtaining valid and reliable reports from teachers is further exacerbated in CAFS, where teacher retention is particularly hard to maintain (Ring and West 2015). While child-reported measures of SEL skills do exist, they have not been established as broadly valid and reliable.

To conduct more rigorous research on the effects SEL programs have on the learning and development of children living in CAFS, we must design robust studies that use assessments that measure the proximal SEL skills of children reliably and validly. It would be especially beneficial if these tools were designed to collect data directly from children in CAFS and allowed users to adapt and contextualize the open-source tools as needed.

### **DEVELOPING A NEW MEASURE OF SEL SKILLS**

In mid-2015, Save the Children was looking for measures to help them understand the effects its growing portfolio of SEL programs were having on primary school-age children in low-resource and crisis contexts. After reviewing the available tools and being faced with the challenges described above, the research team embarked on the process of developing an SEL measure. First, we conducted an internal review to determine which skills we should assess. This review focused not only on Save the Children's programs but also on the competency requirements in national education plans (e.g., Chirwa and Naidoo 2014; Ministry of Education and Sports 2010), guidance on developing SEL standards (e.g., Zinsler, Weissberg, and Dusenbury 2013), and the extant literature on children's SEL. This led us to identify five skills—self-concept, stress management, perseverance, empathy, and conflict resolution. We also identified other skills, such as emotion recognition, emotion regulation, and growth mindset, but we did not include these in the

measurement initiative because they were not an explicit part of Save the Children’s programs for children ages 6-12.

We subsequently reviewed the compendium of SEL measures from high-resource contexts (Denham et al. 2010; Haggerty et al. 2010) to determine possible adaptations and extensions for a new tool. Between 2016 and 2019, we tested several different types and modalities of questions for children—Likert-type scales, visual analog scales, vignettes, and performance-based items—in the Democratic Republic of the Congo, Egypt, Haiti, Iraq, Jordan, Kenya, Mexico, South Sudan, Tanzania, Syria, Thailand, and Uganda (INEE and EASEL 2020). We used three criteria during this iterative process:

- Content validity: Did the skills make sense to program staff and did they think the items satisfactorily captured the dimensions of the skills in their context?
- Internal consistency reliability: Did items measuring the same skill do so in a consistent manner?
- Feasibility: Was the measure easy to adapt and use, and were the results comprehensible for program staff?

As an example, the first version we developed (tested in Mexico and Thailand) included a mix of Likert-type questions and situation-judgment tasks. The Likert-type questions did not have strong internal consistency, and staff members noted that the children struggled with the response categories (“Not at all like me” to “A lot like me”). Indeed, there is evidence that using Likert-type questions may not be appropriate with young children, as they require extensive adaptation, testing, and pretraining (Mellor and Moore 2014; Royeen 1985). The situation-judgment tasks, which were designed to measure empathy, used Selman’s (2003) perspective-taking stage theory to rate children’s ranking of responses to stories of interpersonal conflict. However, program staff members noted that these items were hard to adapt and were not contextually relevant. This iterative design and testing done between 2016 and 2019 led to the development of the ISELA, which describes the development of five SEL skills in children ages 6 to 12: self-concept, stress management, perseverance, empathy, and conflict resolution. The ISELA was designed to be a subscription-free tool that could be adapted and used in program monitoring and evaluation efforts in low-resource and crisis contexts (D’Sa 2019a).

## METHODS

Iraq hosts about 250,000 forcibly displaced Syrians, primarily those who identify as ethnically Kurdish (Durable Solutions Platform 2019). The prolonged conflict in Syria has resulted in children being exposed to severe violence and deprivation, which often has led to social, emotional, and behavioral issues. In a study of displaced Syrian children living in Iraq, about half of the children interviewed told of witnessing the violent deaths of loved ones; 90 percent of these children reported feelings of distress because of this loss (Brophy 2017). The refugee children's caregivers attributed this increase in their fearful and nervous behavior, as well as increased aggression, abuse of substances as a way of coping, and developing speech impediments, on their exposure to violent forced displacement (McDonald et al. 2017). In the latter half of 2018, in response to growing pressure to provide services for displaced children in Iraq, Save the Children started the second phase of a school-based education and child-protection project in Iraq's Dohuk, Ninewa, and Diyala provinces. Our study, which was embedded in the primary schools that were part of this project, focused on establishing the validity and reliability of the ISELA when used with Syrian refugee children living in Iraq. We answered the following research questions:

- Construct validity: How well are the observed variables predicted by the underlying SEL constructs?
- Criterion validity: To what extent are the SEL skills associated with variables that we theorize they should have a strong relationship with?
- Internal consistency reliability: How well do the items measuring each SEL skill relate to other items measuring that skill?
- Interrater reliability: What is the level of agreement in scoring items between different assessors?

## RESEARCH DESIGN

Initial discussions with the program staff revealed that most of the Syrian refugee children who were to participate in this study were fluent in Syrian Kurdish but not Arabic. The Syrian Kurdish dialect these children spoke often was not written; even the youth and adult assessors had a hard time reading the script. Hence, we decided to translate the ISELA into colloquial Mesopotamian Arabic and to have 20 bilingual assessors administer the tool to children in Syrian Kurdish.

This in situ translation could have biased or altered responses if the assessors asked questions using different wording, so to limit such variations, the assessors debated the Kurdish translation during the training and reached a consensus on wording. During the data collection in February 2019, a trained team leader supervised each five-person data-collection team. This study was approved by the ethics review committee at Save the Children. The interviewers obtained informed consent from caregivers and assent from the children before data were collected.

SAMPLE

This study focused on the ten schools in Dohuk that only served Syrian refugee children. We used a root mean square error of approximation (RMSEA) sample size calculator (MacCallum, Browne, and Sugawara 1996; Preacher and Coffman 2006) to estimate the required sample, given the degrees of freedom for each level-one and level-two factor analysis model. Based on our assumptions ( $\alpha=.05$ ,  $\text{power}=.80$ ,  $\text{RMSEA for alternative distribution}=.05$ , and  $\text{RMSEA for null distribution}=.02$ ), we needed to collect data from a minimum of 500 children. In each of the ten schools, eight boys and eight girls were randomly sampled from each classroom for grades 1-4. After removing children who did not consent and because of the small class sizes in some schools (where enumerators assessed all children), we ended up with a sample of 620 children. In Table 1 we illustrate the demographic characteristics of our sample.

Table 1: Demographic Characteristics of Syrian Refugee Children from Dohuk, Iraq, Included in the ISELA Psychometric Pilot

Demographics	n=620
Female	49%
Age (average years)	9
Grade 1	26%
Grade 2	25%
Grade 3	25%
Grade 4	24%
Kurdish mother tongue	92%
Arabic mother tongue	13%
Multilingual	5%
SES (average of 9 household items)	6
Risks (average of 7)	1.2
Interpersonal threats (average of 10)	3.5

## INSTRUMENT

The ISELA measures five SEL skills—self-concept, stress management, perseverance, empathy, and conflict resolution—with information drawn from six subtasks.

**Self-concept** is children's beliefs in their skills and abilities to meet present and future goals (Bandura 2006). In the first version of the ISELA (tested in 2016), we created a Likert-type questionnaire to measure children's growing capacity for independence and confidence in their daily activities, using Bandura's (2006) guidance for developing self-concept measures. However, these Likert-type items demonstrated weak internal consistency. Moreover, colleagues working with refugee children in Thailand and Jordan noted that their work was focused on helping children have a future vision for themselves because that was a significant challenge for the children they worked with. Given the vast literature on children's general and domain-specific self-concept and self-efficacy (Bong and Skaalvik 2003), we grounded the ISELA's measurement of self-concept on children's future orientation (Markus and Nurius 1986); that is, their vision of the “selves we imagine ourselves becoming in the future, the selves we hope to become, the selves we are afraid we may become, and the selves we fully expect we will become” (Oyserman and Fryberg 2006, 4). During the administration of the ISELA, children are asked to draw and reflect on two future selves and to identify a barrier to and support for each (six items: J3-J8). Assessors' scores reflect whether the participants can imagine a future and articulate what would support them or stop them from attaining that future.

**Stress management** is the conscious use of personal skills and resources to reduce the effects of chronic stress and/or acute adversity (Rutter 1981). Since most of Save the Children's SEL programs focused on teaching children explicit strategies to manage stress (like belly breathing or counting to ten), the subtask (three items: D1-D3) asks children to identify strategies they use to calm down if angry or upset. Assessors' scores reflect whether the child can accurately define up to three appropriate, nondestructive stress-management strategies. If the child cannot identify one appropriate strategy, they are not asked for additional strategies.

**Perseverance** refers to a child's ability to stay on task despite setbacks and the task being hard to complete (Von Culin, Tsukayama, and Duckworth 2014). In the context of SEL skills, it refers to the child's ability to motivate himself/herself to continue engaging in hard and complex social relationships. The perseverance subtask on the ISELA (four items: F1-F4) asks children to draw four increasingly

difficult geometric figures using their nondominant hand. At 20 and 40 seconds into each of the four drawing activities, the assessor asks the child if they want to stop and move on to the next subtask. Assessors' scores reflect not the accuracy of the child's drawing but their ability to persist with each drawing for 60 seconds; if a child asks to stop, they are not shown subsequent geometric figures.

**Empathy** is the ability to consider the perspective of other individuals, understand their emotional reactions, and integrate that into socially desirable interactions (Selman 2003). The empathy subtask (10 items: E1-E10) first asks participants to recognize sadness (E1) and anger (E6) in pictures of two children. E1 and E6 were adapted from the Assessment of Children's Emotion Skills (Schultz, Izard, and Bear 2004). The participants are then asked to describe four things (E2, E3, E7, E8) they could do to make the sad/angry child feel better—an adaptation of the International Development and Early Learning Assessment (Pisani, Borisova, and Dowd 2018)—before being asked to interpret the intentions of the child whose action (e.g., pushing or spilling water) caused the original child to feel sad/angry. The last four items (E4, E5, E9, E10) were developed to measure hostile attributional bias, which is children's tendency to attribute hostile intent to ambiguous provocations. We theorize that these four items measure conflict resolution, since children who attribute hostile intent to ambiguous provocations tend to negotiate aggressive resolutions to interpersonal conflict (Dodge et al. 2015).

**Conflict resolution** refers to the strategies and methods children use to peacefully negotiate interpersonal disputes with their peers (Lemerise and Arsenio 2000). Conflict resolution is assessed through children's interpretation of an interpersonal conflict vignette (four items: G1-G4), along with the four items from the empathy subtask. Participants are asked to name two things they would do if they were playing with a toy and another child asked to play with the same toy. They are then asked to name two things they would do if the other child took the toy from them without asking and started playing with it. These conflict-resolution subtask items were adapted from the International Development and Early Learning Assessment (Pisani et al. 2018) and the Challenging Situations Task (Denham et al. 2013).

**Relationships** focus on the individuals who are part of the child's social network. The relationship subtask does not measure a specific SEL skill, but items from this subtask are theorized to load onto the measurement of four of the five SEL skills described above. During prior conversations about the ISELA's content validity, program colleagues noted that the items from the SEL skill subtasks (described above) did not capture the skill adequately. Adding items from the

relationship subtask enabled us to broaden the measurable dimensions of these skills. In the relationship's subtask, participants are asked to describe the size of their social network, including family, peers, and community adults. For each type of individual in their social network, participants are asked if they talk to someone when they are sad (stress management), can ask for help when working on something difficult (perseverance), know when a person in their social network is feeling sad (empathy), and can ask for help with a problem with a peer (conflict resolution).

### **ADAPTATION OF THE ISELA**

All questions in the ISELA have a binary response option: correct/incorrect or appropriate/inappropriate. Participants provide open-ended responses, and the assessors mark the responses as appropriate or inappropriate while collecting the data. This in situ coding was done for two reasons. First, it improved the feasibility of collecting data. In early versions of the ISELA, we collected actual participant responses for some questions and coded them post hoc. However, by investing in assessor training, we were able to reduce the time and resources needed to score the assessment, thereby improving feasibility. Second, given the normative nature of SEL skills, the open-ended response coding improves the adaptability of the ISELA. In each new context, the research team develops a list of socially and contextually appropriate and inappropriate response options. In Iraq, the program team from Save the Children developed the first list of appropriate and inappropriate response options for the stress-management, empathy, and conflict-resolution subtasks. The assessors, who were adults from the same community as the children, added to this list during training. After a one-day field test with children (which was not included in the final sample), the assessors refined the list.

Given the high rate of exposure to adversity among forcibly displaced Syrian children, we included questions about the following eight risk factors:

1. Have you ever had to work to earn money to support your family?
2. Have you ever had to miss school for longer than one month?
3. Have you ever had to leave your home because it was not safe?
4. Have you ever lived in a home where an adult regularly did drugs?
5. Have you ever lived in a home where people shouted or yelled?

6. Have you ever lived in a home where people pushed, slapped, or threw something?
7. Have you ever gone hungry because there was not enough food?
8. Have you ever had a family member leave home for more than six months?

Next, to capture the nature of the interpersonal threats they may have experienced, we asked children whether any of the following had occurred at their school in the previous week:

1. Did you feel afraid?
2. Did you feel afraid on your way to school?
3. Did a child tease another child?
4. Did a child leave out another child?
5. Did anyone say something mean?
6. Did children get into a fight?
7. Did anyone throw something to hurt another?
8. Did an adult scream or yell?
9. Did an adult hit or kick anyone?
10. Did an adult threaten to hurt a child?

## **DATA ANALYSIS**

### **FACTOR ANALYSIS**

To validate the structure of the five SEL constructs—self-concept, stress management, perseverance, empathy, and conflict resolution—we first fit an exploratory factor analysis (EFA) model (in Mplus 7.4) using a geomin rotation on a random half of the data, specifying that all observed variables were categorical. After observing the eigenvalues and fit statistics for the different EFA models

(e.g., one factor, two factor), we chose the model that had the strongest fit. We confirmed our EFA on the other random half of the data, fitting a confirmatory factor analysis model for each construct separately (weighted least squares means and variance estimator appropriate for categorical data). We observed the fit statistics to see how well the specified model fit the data (Schreiber et al. 2006).

### CONVERGENT AND DISCRIMINANT VALIDITY

We used grade, experience of interpersonal threats, and exposure to risk factors to establish convergent and discriminant validity. We used grade instead of age because 10 percent of the children did not provide their age, and age and grade were strongly correlated ( $r=0.75$ ). We hypothesized that grade would have a statistically and practically meaningful positive relationship with the five SEL skills. We also hypothesized that children with higher empathy and conflict-resolution scores would have the interpersonal skills to identify more interpersonal threats at their school. This hypothesis was based on Lemerise and Arsenio's (2000) integrated model of social information processing. They explain that the "intensity with which children experience emotions and their skill at regulating emotion will influence what is noticed and the meaning attributed to the situation" (113). Children with stronger SEL skills likely notice more of the interpersonal threats around them and encode these experiences in their memory to use when processing future social cues. Alternatively, we expected the risk factors the children faced to be negatively associated with their SEL skills (Anda et al. 2011; Betancourt et al. 2013).

To understand this convergent and discriminant validity, we fit a multilevel model for each SEL skill with grade, index of interpersonal threats, and index of risk factors as our predictors. We controlled for the differential effects of gender and household wealth and clustered the standard errors at the school level. The items we use in the ISELA artificially censor children's responses at the higher and lower ends of the scale. For example, in the empathy subtask, we asked children to recognize sadness/anger and what they would do to help another child feel better. The developmental trajectory of empathy likely extends to more foundational emotion-recognition skills, as well as more advanced third-order perspective-taking. Since the subtasks in the ISELA cannot capture this entire developmental trajectory due to time and resource constraints, we may be artificially censoring

children at the lower and higher ends of the scale for each skill. To deal with this censoring, we used a Tobit regression in fitting the multilevel models. Tobit blends ordinary least squares with a probit function to deal with the censoring of data (Stewart 2013).

### **INTERNAL CONSISTENCY RELIABILITY**

We used the Kuder-Richardson-20 (KR-20) statistic, a variation of the Cronbach's alpha statistic designed specifically for use with dichotomous items, to understand how reliably the individual items under each ISELA construct measure that skill. The KR-20 statistic ranges from 0 to 1; values between 0.70 and 0.95 suggest strong internal consistency (Streiner 2003).

### **INTERRATER RELIABILITY**

For the first boy and first girl from each grade assessed at each school (n=99), the assessors worked in pairs, with one assessor conducting the assessment and scoring while the other listened and scored. These data were analyzed using Krippendorff's alpha statistic, with values of 0.8-1 representing a strong level of agreement between the assessors (McHugh 2012).

## **RESULTS**

### **FACTOR ANALYSIS**

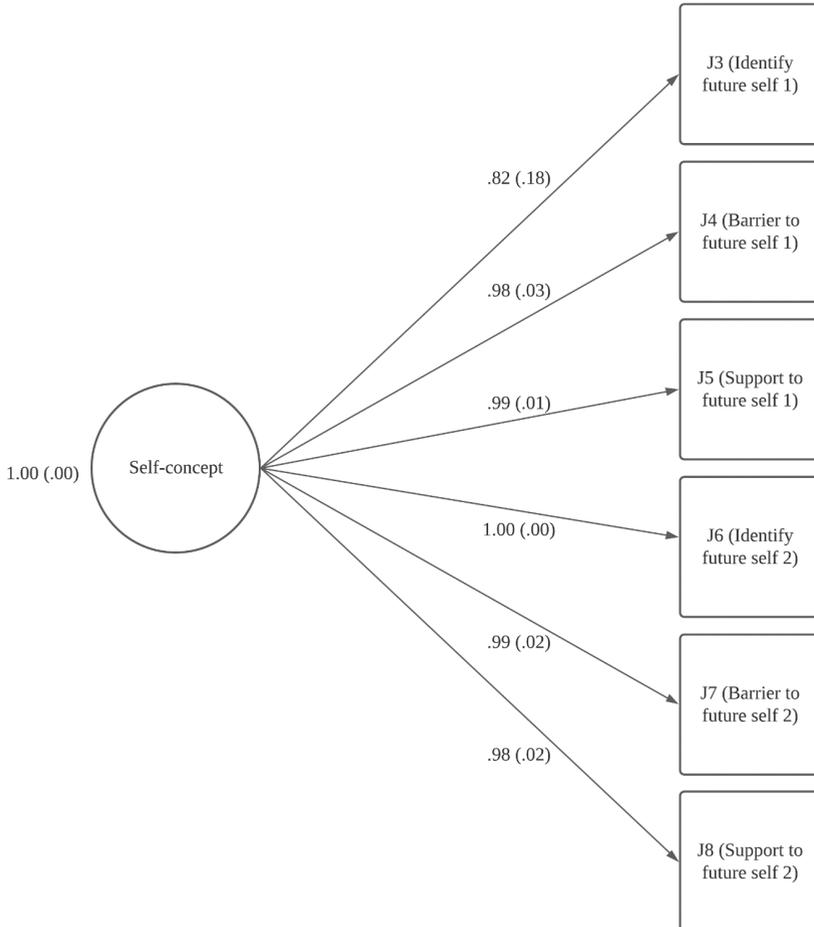
We fit a separate measurement model for each SEL skill. In Table 2 we provide an overview of the statistics we used to judge how well each model fit the data. In all five models, the comparative fit index and Tucker-Lewis index were above our acceptance threshold of 0.95 (Schreiber et al. 2006), which suggests that the models fit the data well when compared to a baseline model where all the paths were uncorrelated. The absolute fit indices—root mean square error of approximation and standardized root mean square residual—for all measurement models were lower than the prespecified thresholds of 0.06 and 0.08 (Schreiber et al. 2006), which suggests that our models fit the data well. Below we provide illustrations for each of the five measurement models and describe the structure of the loadings for each latent construct.

*Table 2: Goodness of Fit Statistics from Confirmatory Factor Analysis Models Fit with Random Half of Data from Iraq*

		Self-Concept	Stress Management	Perseverance	Empathy	Conflict Resolution
Split-half sample	n	309	309	309	309	309
Observed variables	x	6	4	3	5	6
Chi-square test of model fit	$\chi^2$ (df, p-value)	8.93 (9, 0.44)	0.32 (2, 0.85)	0 (0, <0.001)	1.95 (4, 0.75)	11.93 (7, 0.10)
Comparative fit index	CFI	1	1	1	1	0.99
Tucker-Lewis index	TLI	1	1	1	1	0.97
Root mean square error of approximation	RMSEA (CI)	<0.001 (0-0.06)	<0.001 (0-0.06)	<0.001 (0-0)	<0.001 (0-0.06)	0.05 (0-0.09)
Standardized root mean square residual	SRMR	0.03	0.01	<0.001	0.02	0.06
Weighted root mean square residual	WRMR	0.22	0.12	0.01	0.23	0.56

**Self-concept.** In Figure 1, we illustrate the loading of the latent self-concept factor on the six observed variables from the self-concept subtask. We did not modify the model or any of its paths. We found that the latent self-concept construct predicted the six observed variables with almost equal loadings on all predicted pathways.

Figure 1: Standardized Loadings for Measurement Model for Latent Self-Concept Construct

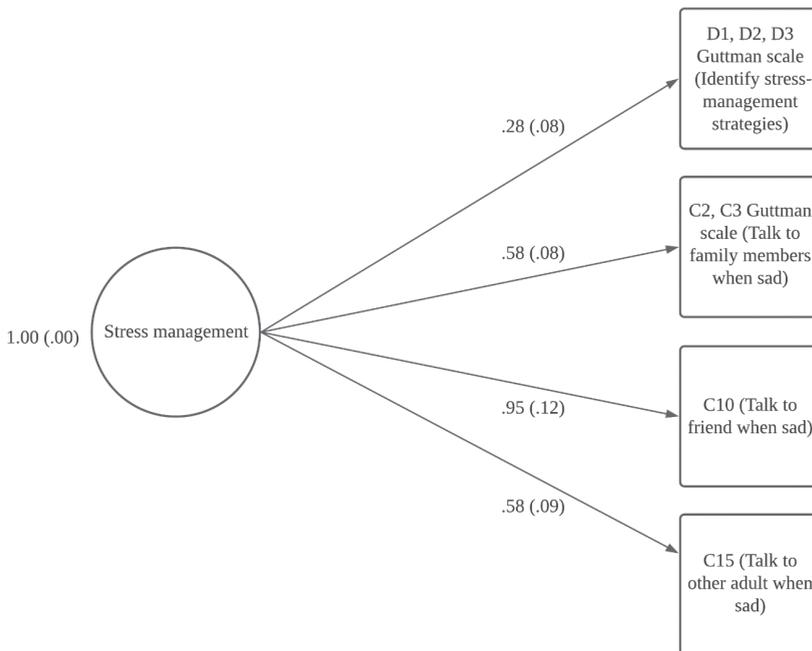


Note: All illustrated paths are statistically significant at  $p < 0.05$ .

**Stress management.** We hypothesized a single latent factor predicting items from the stress-management subtask and the relationship subtask (participants reach out to their social network when sad). We formed a Guttman scale (summed scale ranging from 0 to 3) with the three stress-management subtask questions because they were hierarchical: participants unable to identify an appropriate stress-

management strategy were not asked for subsequent strategies. Since we included these items in the same measurement model with other items that were from a different subtask and had a different response structure, we had to account for the hierarchical dependency. We also created a summed scale (range 0-2) with the first two items from the relationship questions, since participants who could not identify a family member to talk to when sad were not asked about a second family member. In our final model (see Figure 2), the stress-management latent construct predicted the four observed variables; the model fit the data well and was aligned with our initial hypothesis. The latent factor most strongly predicted participants' response to whether they talked to a friend when sad.

Figure 2: Standardized Loadings for Measurement Model for Latent Stress-Management Construct

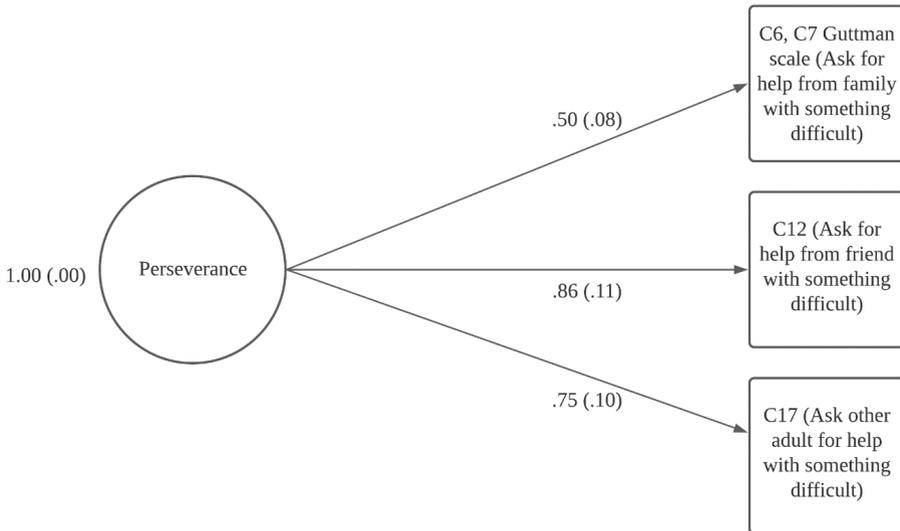


Note: All illustrated paths are statistically significant at  $p < 0.05$ .

**Perseverance.** We hypothesized that a single latent factor would predict five observed variables. First was a summed scale (range 0-4) of questions drawn from the perseverance subtask. Second was a summed scale (range 0-2) about which family members participants sought help from when working on something difficult. The final three dichotomous items included two relationship questions: whether the child approached a friend or community adult for help when working on something difficult, and whether the child chose to do the self-concept drawing.

The EFA model suggested that there were two latent factors, one predicting the three relationship items and the other predicting the drawing items from the perseverance and self-concept subtasks. We were unable to confirm this two-factor model, since one of the factors would have two observed variables and the model would not be identified. Additionally, the loadings for the two drawing items from the EFA were negative and did not fit our theory. Hence, we fit a single latent perseverance factor, predicted by the three observed relationship variables (see Figure 3). This model was fully saturated and precluded interpretation of the model fit. However, all factor loadings were strong, positive, and significant.

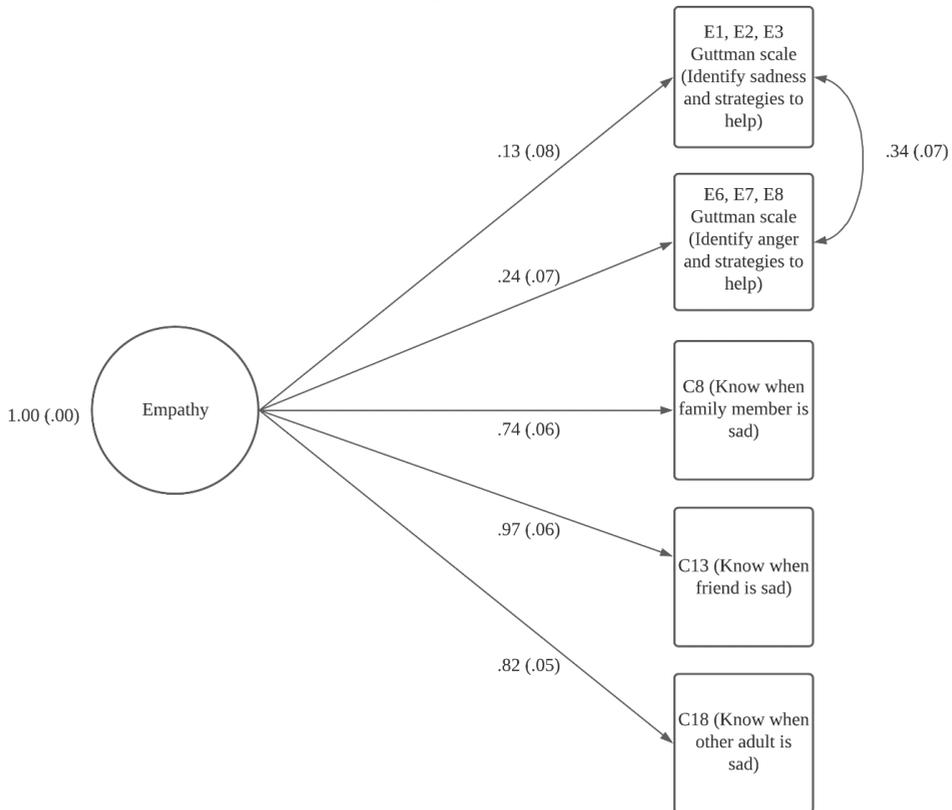
Figure 3: Standardized Loadings for Measurement Model for Latent Perseverance Construct



Note: All illustrated paths are statistically significant at  $p < 0.05$ .

**Empathy.** We theorized a single latent factor that predicted five observed variables. The first two variables were summed scales (range 0-3) from the empathy subtask: (1) recognize sadness and identify two things to do to help a sad child feel better, and (2) recognize anger and identify two things to help an angry child calm down. The last three items were from the relationship subtask: recognize when a family member, friend, or community member is sad. The final model (see Figure 4) fit the data well and was aligned with our initial theory. The one modification we made was to allow the residual variances on the two summed scales from the empathy subtask to covary. The latent factor most strongly predicted participants' response to whether or not they could identify when someone in their social network was sad.

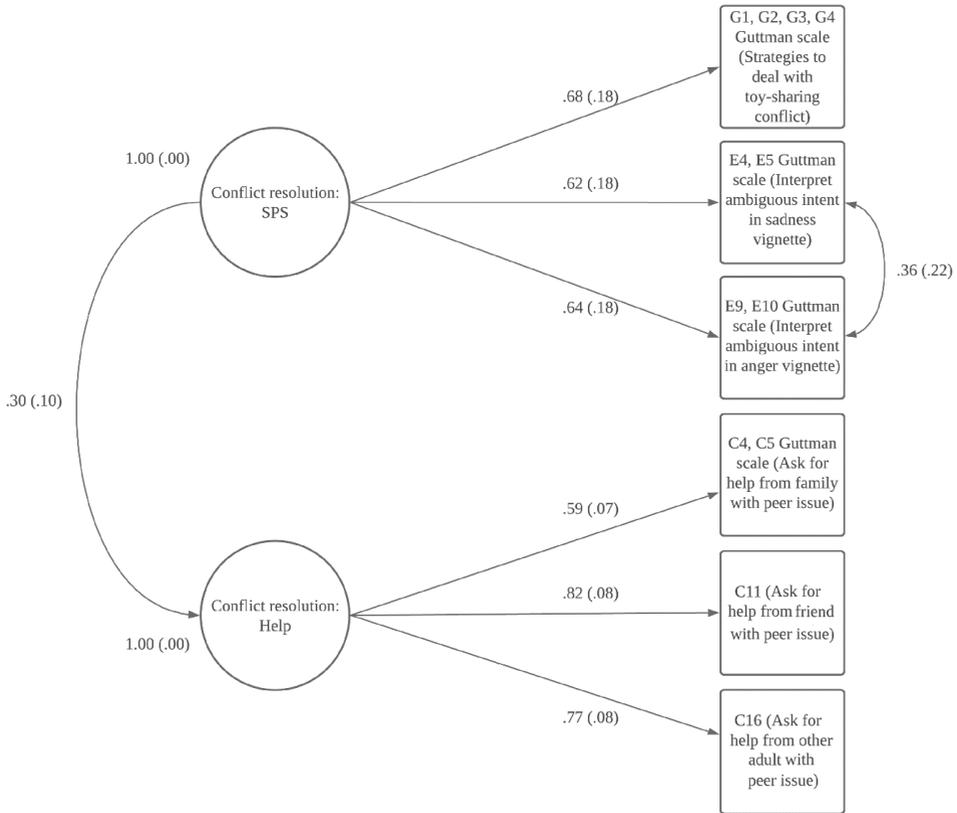
Figure 4: Standardized Loadings for Measurement Model for Latent Empathy Construct



Note: All illustrated paths are statistically significant at  $p < 0.05$ .

**Conflict resolution.** For conflict resolution, we theorized a two-factor model. The first latent factor—social problem-solving (SPS)—would be predicted by three summed-scale variables. The first (range 0-4) was a summation of the four items from the conflict-resolution subtask. The second and third were summations (range 0-2) from the empathy subtask: whether the participant attributed benign intent to ambiguous provocations in the sadness and the anger vignettes. The second latent factor—help—would be predicted by three observed variables from the relationship subtask: whether participants asked a family member, friend, or community member for help when resolving a peer issue. We expected the two latent constructs—SPS and help—to have a strong and statistically significant correlation, since they were measuring facets of conflict resolution. We made one modification to allow the residuals of the two summed variables from the empathy subtask to covary. This modified model (see Figure 5) fit the data well and all the paths were statistically and practically meaningful. The two latent factors—SPS and help—had a strong and positive correlation of 0.30.

Figure 5: Standardized Loadings for Measurement and Structural Model for Latent Conflict-Resolution Construct



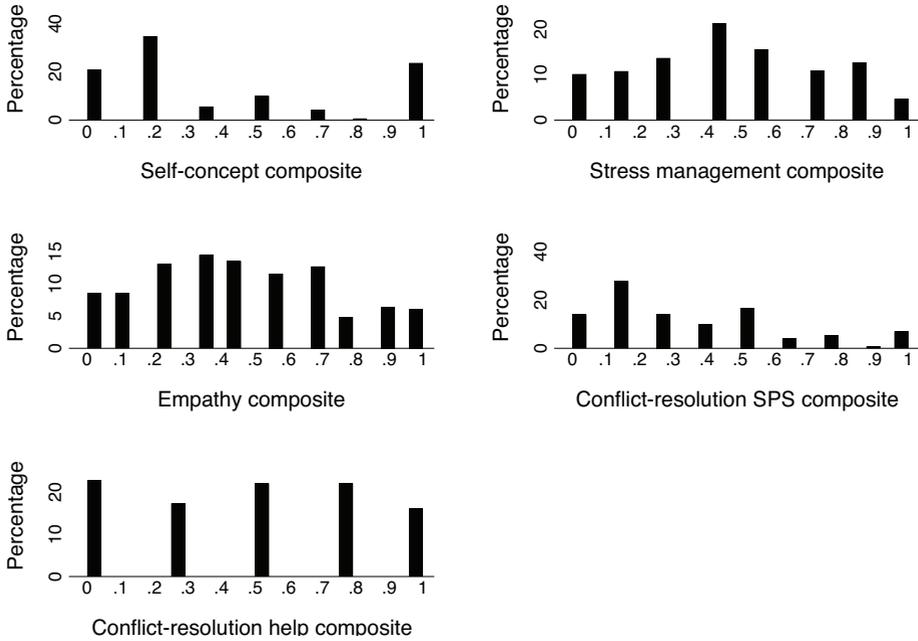
Note: All illustrated paths are statistically significant at  $p < 0.05$ .

### CREATING COMPOSITE SCORES

Because the perseverance confirmatory factor analysis model was fully saturated, we were unable to interpret its fit and validate its structure; thus, we did not have sufficient evidence to proceed with further testing of the validity and reliability of this construct. Having established the validity of the internal structure of the other four SEL skills, the next step was to create composite scores to use in further analysis. One method was to use the confirmatory factor analysis loadings as weights for each item. However, given that the ISELA is used by teams in CAFS with limited analytic resources, we followed the more commonly applied research practice: create a sum for each skill and divide it by the total number of items. In so doing, we assume that all items are equally weighted in the composite score. We found a very strong and positive correlation between the unit-weighted and loading-weighted composites, which ranged from 0.83 for empathy to 0.99 for self-concept.

In Figure 6, we present five distribution histograms. The distributions for stress management and empathy were close to normal. The self-concept and help constructs had skewed distributions, with approximately 45 percent of participants scoring at the two extremes.

Figure 6: Distribution of Composite Scores for the SEL Skills



The five constructs had moderately strong, positive correlations with each other (see Table 3), ranging from 0.09 between self-concept and help to 0.53 between empathy and SPS. Overall, the correlations suggest that, while these constructs are related to each other, they still help capture different skills in children’s social and emotional development.

*Table 3: Correlations of the Five Unit-Weighted SEL Composites*

	Self-Concept	Stress Management	Empathy	Conflict Resolution: SPS	Conflict Resolution: Help
Self-concept	1				
Stress management	0.21	1			
Empathy	0.26	0.47	1		
Conflict resolution: SPS	0.35	0.35	0.53	1	
Conflict resolution: Help	0.09	0.47	0.27	0.28	1

### CONVERGENT AND DISCRIMINANT VALIDITY

In Table 4 we present the results of fitting the five Tobit models to establish the convergent validity with grade and interpersonal threats, and the discriminant validity with exposure to risk factors. Our sample size was 576 children, due to missing data across the covariates in the model.

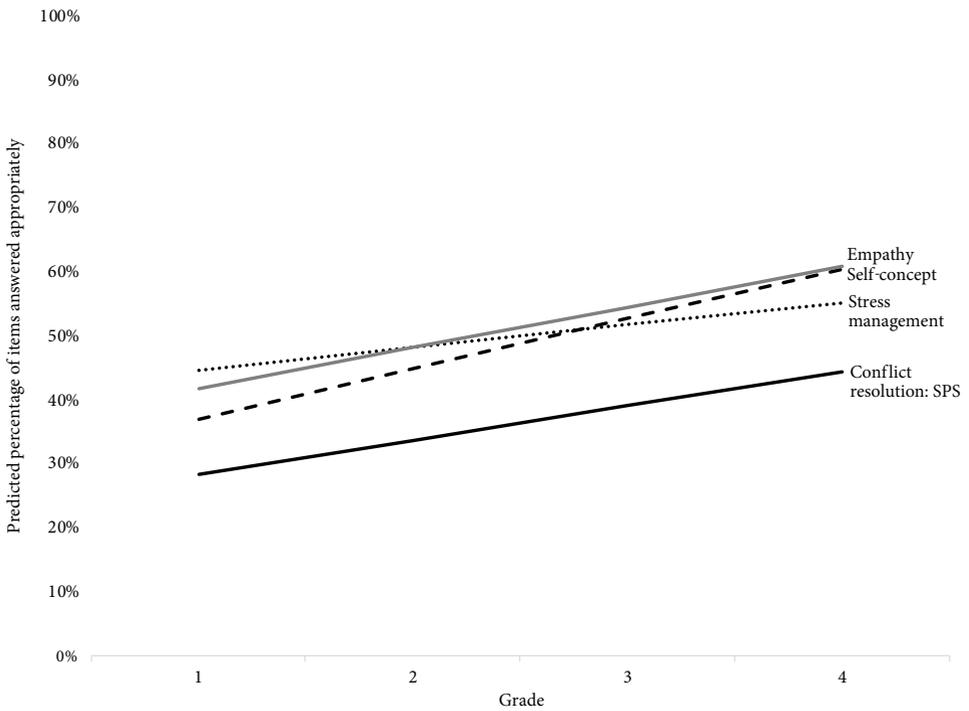
Table 4: Tobit Model Estimates for Predicted Social and Emotional Skill Scores

	Self-Concept	Stress Management	Empathy	Conflict Resolution: SPS	Conflict Resolution: Help
Grade	0.0779** [0.03]	0.0354* [0.02]	0.0635*** [0.01]	0.0532** [0.02]	0.0526 [0.03]
Gender	0.00625 [0.05]	-0.0468 [0.03]	-0.0403~ [0.02]	0.0114 [0.03]	0.00846 [0.04]
Socio-economic status	-0.173 [0.31]	0.104 [0.10]	0.230~ [0.14]	0.227* [0.11]	0.422 [0.30]
Number of risk factors	0.00186 [0.02]	-0.0111 [0.02]	0.00714 [0.02]	-0.0239 [0.02]	-0.0146 [0.02]
Number of threats	0.0228 [0.02]	0.0294*** [0.00]	0.0312*** [0.01]	0.0246** [0.01]	0.0454*** [0.01]
Intercept	0.321 [0.30]	0.276** [0.09]	0.105 [0.12]	0.0192 [0.12]	-0.0413 [0.22]
Sigma	0.612*** [0.08]	0.310*** [0.01]	0.305*** [0.02]	0.322*** [0.04]	0.537*** [0.03]
df	5	5	5	5	5
n	576	576	576	576	576

Note: Standard errors clustered at the school level. Standard errors in brackets.  
p<0.10, \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

As theorized, we found a statistically and practically meaningful positive relationship between grade and four of the SEL skills; the child's grade was not predictive of their score on the help construct. However, for the other four constructs, a one-grade difference was associated with a 4- to 8-percentage-point change in the number of items children answered appropriately (see Figure 7).

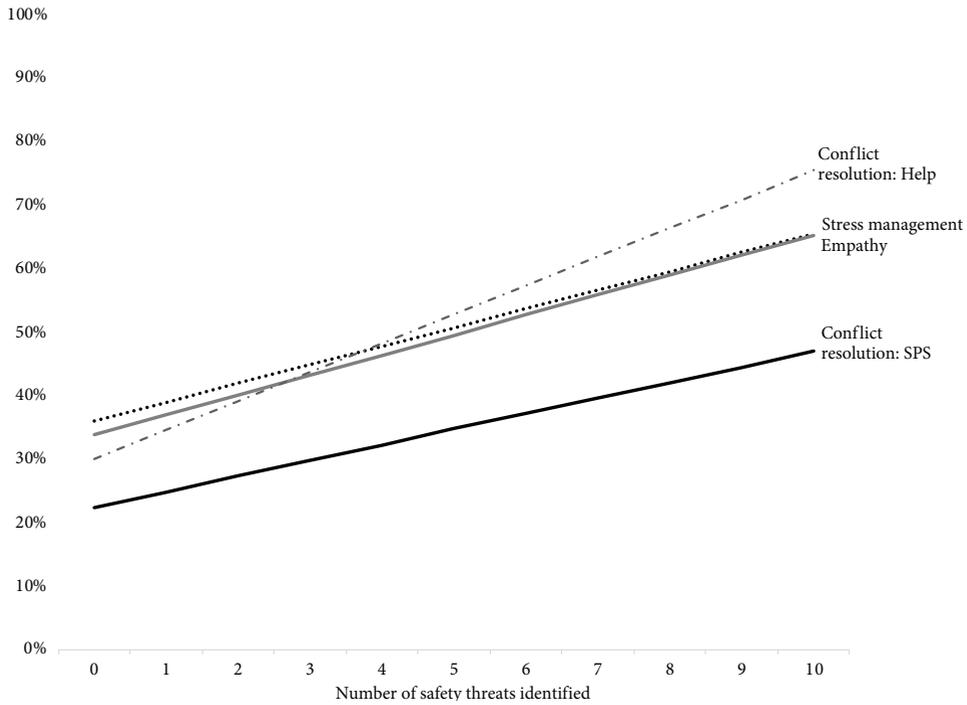
*Figure 7: Predicted Social and Emotional Skill Scores by Grade, Controlling for Gender, Socioeconomic Status, Exposure to Risk Factor, and Interpersonal Threats*



Note: n=576, standard errors clustered at the school level

Additionally, as initially theorized, children’s self-reported perception of interpersonal threats was positively associated with their empathy and conflict-resolution scores. We found one relationship that we had not initially theorized: children’s self-reported perception of interpersonal threats was positively associated with their stress-management scores. A one-unit change in the number of threats a child identified was positively associated with a 2- to 5-percentage-point change in their SEL skills (see Figure 8).

Figure 8: Predicted Social and Emotional Skill Scores by Number of Identified Interpersonal Threats, Controlling for Gender, Grade, Socioeconomic Status, and Exposure to Risk Factors



Note: n=576, standard errors clustered at the school level

Finally, we predicted that there would be a negative relationship between children’s SEL skills and exposure to risk factors. However, we found no relationships between these variables when controlling for the effect of important equity factors, such as grade, gender, and household wealth.

### INTERNAL CONSISTENCY RELIABILITY

Across the five SEL skills, we found good internal consistency reliability of 0.70 or higher (see Table 5). This suggests that the individual items within each SEL construct are strongly correlated and measure a similar skill.

*Table 5: Internal Consistency Reliability (KR-20) for SEL Constructs*

Social and Emotional Competency	KR-20
Self-concept	0.91
Stress management	0.70
Empathy	0.77
Conflict resolution: SPS	0.82
Conflict resolution: Help	0.70

**INTERRATER RELIABILITY**

The Krippendorff’s alpha statistic for each observed variable included in the five SEL constructs was strong, above our predetermined acceptance threshold of 0.80 (see Table 6). This suggests that different assessors were administering the ISELA consistently and reliably during the data collection.

*Table 6: Krippendorff’s Alpha Interrater Reliability Statistics for Observed Variables Included in Each of the SEL Constructs*

Social and Emotional Competency	Krippendorff’s Alpha Range for Observed Variables
Self-concept	0.86-0.98
Stress management	0.96-1
Empathy	0.89-1
Conflict resolution: SPS	0.94-1
Conflict resolution: Help	0.86-0.96

**DISCUSSION**

The proliferation of SEL programs in CAFS necessitates a fuller understanding of the impact these programs have on children. However, attempts to build this knowledge have been stymied by the dearth of valid, reliable, and feasible measures that are open source and subscription free. In this article, we provide strong, positive evidence for the construct validity, internal consistency reliability, and interrater reliability of the ISELA when used with Syrian refugee children living in Iraq. We also have promising preliminary evidence for the criterion validity of the SEL skills we measured. However, designing a measure for use in CAFS comes with several tradeoffs related to assessment feasibility versus psychometric rigor, each tradeoff having a different degree of impact on the usability or rigor of the tool. Because of the resource and logistical limitations of collecting data in nonpermissive, politically volatile, or geographically challenging locations in

CAFS (Anderson, Read, and Losada 2020; Halman 2019), we need assessments that are brief, easy to use, and require limited materials or stimuli. Below we discuss the findings from this study in light of the feasibility-psychometric rigor tradeoffs we made and focus on implications for future SEL measurement in CAFS.

A challenge of measuring SEL skills is to ensure that we capture the dimensions of individual skills (Halle and Darling-Churchill 2016). For example, children's judgment about interpersonal conflict resolution is affected by several factors, including the intent they attribute to the other child, their interpersonal resolution strategies, and their help-seeking behavior (Lemerise and Arsenio 2000; Selman 2003). We attempted to capture different skill dimensions in the ISELA by ensuring that items loading on each skill construct came not only from the skill subtasks but also from the relationship subtask. This decision came with an inherent tradeoff. We increased the assessment time (about 25 minutes per child), a decision that meant we had to make cuts to other parts of the assessment. Nonetheless, we were able to establish content validity with partners who were using the tool, as well as the factor structure of self-concept, stress management, empathy, and conflict resolution.

However, we were unable to validate the factor structure for perseverance. One issue with the perseverance subtask was that 79 percent of participants completed all four drawings. This meant that there was little variation in participants' performance. This was a symptom of the tradeoff we made: by adding the relationship subtask, we shortened other ISELA subtasks; we wanted a short, performance-based measure of perseverance that was easy to use with few stimuli and materials. To measure perseverance most precisely, we need to include more complex and time-intensive items that capture the diversity of children's capacity for perseverance. In future iterations of the ISELA, we intend to work on further validating the factor structure of the perseverance latent construct by testing adaptations, including harder geometric figures and increased task complexity (e.g., not being able to lift the pencil off the paper).

The assessment-brevity tradeoff affected all sections of the ISELA, which resulted in our being judicious about the number of subtask items and the complexity of child response options. This raised a statistical challenge when trying to ensure that different items measuring each skill demonstrated strong internal consistency. All of the variables in the ISELA have a dichotomous construction and the number of items measuring each skill is small, between 4 and 12. This was done to make the assessment brief and easy to use in CAFS by limiting the need to explain complex response options to the children. Variables with larger scoring ranges and constructs with more items from the same subset of the survey provide stronger

internal consistency reliabilities (Bentler 2009; Streiner 2003). However, our analysis demonstrated that, even with these limitations, the items measuring each of the SEL skills in the ISELA demonstrated strong internal consistency reliability. Our findings suggest that measuring children's SEL skills in CAFS does not require sacrificing a measure's internal consistency. An iterative development process, like that we followed with the ISELA, can result in measures that are brief and reliable.

Another tradeoff was to remove the off-site scoring of children's responses to vignettes and scenarios; instead, we created an in situ protocol for assessors to score children's responses as socially and contextually appropriate or inappropriate. These interpretive response categories add to the ISELA's adaptability. This measure adaptability is important in CAFS because severe adversity can affect children's normative pathways of development (Betancourt et al. 2013, 2017), which requires using measures that can be changed quickly and easily to meet contextual needs. However, the adaptability of in situ scoring adds structured bias on the part of the assessor. Nonetheless, we found strong interrater reliability for all the ISELA items, which suggests that assessors had strong levels of agreement when scoring the same child. Generating in-depth response options and rigorous training can prepare assessors to score children reliably. In Iraq, we trained the assessors for three full days and conducted a field test on day four. We would recommend a similar investment in assessor training for future uses of the ISELA.

The final tradeoff we made was not with the ISELA but with the process of testing it in Iraq. Because of time and resource constraints, we were unable to include a validated measure of the children's academic or psychosocial development to establish criterion validity. This limited the scope and depth of the validity testing we could accomplish. We instead used grade (as a proxy for age because of missing data) and exposure to interpersonal threats to establish convergent validity. While we were able to provide promising evidence for the convergent validity of self-concept, stress management, empathy, and conflict resolution, these results are preliminary and need further investigation. Additionally, we were unable to establish the discriminant validity of the SEL skills with an index of distal risk factors. Unlike the extant literature that has included proximal risk factors (e.g., witnessing torture) in studies of forcibly displaced children's exposure to trauma (Morgos et al. 2007; Smith et al. 2002), we decided not to include risk factors directly related to the Syrian refugee children's experiences of trauma. We wanted to have a more robust referral system in place before asking young children sensitive questions about their displacement. Also, the floor effects observed in the number of risk factors the children experienced (the average child reported 1.2 of 7 risk factors) suggests social desirability bias or that the distal risk factors were

not pertinent. Future ISELA validation efforts should include more pertinent and robust measures of academic and psychosocial development to establish criterion validity. For example, the Inter-Agency Network for Education in Emergencies' Measurement Library (on which the ISELA is available) is compiling valid, reliable learning and development measures that can be used in CAFS. Using a measure from this initiative can further our understanding of the validity of the ISELA.

## CONCLUSION

Between 2016 and 2019, we embarked on an iterative design and testing process to develop a psychometrically rigorous yet feasible open-source measure of children's SEL skills. The resulting tool—the ISELA—was specifically designed with the resource and logistic constraints of CAFS in mind. The current study provides strong evidence that the ISELA measures the self-concept, stress-management, empathy, and conflict-resolution skills of Syrian refugee children living in Iraq validly and reliably. To expand the evidence based on the impact SEL programs have on children, the tool can be used reliably by groups of assessors in the context of skills-monitoring or impact evaluations in CAFS.

## REFERENCES

- Aber, J. Lawrence, Catalina Torrente, Leighann Starkey, Brian Johnston, Edward Seidman, Peter Halpin, Anjuli Shivshanker, Nina Weisenhorn, Jeannie Annan, and Sharon Wolf. 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): 507-29. <https://doi.org/10.1080/19345747.2016.1236160>.
- Anda, Robert F., Vincent J. Felitti, J. Douglas Bremner, John D. Walker, Charles Whitfield, Bruce D. Perry, Shanta R. Dube, and Wayne H. Giles. 2011. "The Enduring Effects of Abuse and Related Adverse Experiences in Childhood: A Convergence of Evidence from Neurobiology and Epidemiology." *European Archives of Psychiatry and Clinical Neuroscience* 256 (3): 174-86. <https://doi.org/10.1007/s00406-005-0624-4>.
- Anderson, Kate, Lindsay Read, and Elena Losada. 2020. *Academic Learning Measurement and Assessment Tools in Education in Emergencies: Identifying, Analyzing, and Mapping Tools to Global Guidance Documents*. New York: INEE. <https://inee.org/resources/academic-learning-measurement-and-assessment-tools-education-emergencies-identifying>.

- Bandura, Albert. 2006. "Guide for Constructing Self-Efficacy Scales." In *Self-Efficacy Beliefs of Adolescents*, edited by Frank Pajares and Tim Urdan, 307-37. Greenwich, CT: Information Age Publishing.
- Bentler, Peter M. 2009. "Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability." *Psychometrika* 74 (1): 137-43. <https://doi.org/10.1007/s11336-008-9100-1>.
- Betancourt, Theresa S., Ryan Mcbain, Elizabeth A. Newnham, and Robert T. Brennan. 2013. "Trajectories of Internalizing Problems in War-Affected Sierra Leonean Youth: Examining Conflict and Postconflict Factors." *Child Development* 84 (2): 455-70. <https://doi.org/10.1111/j.1467-8624.2012.01861.x>.
- Betancourt, Theresa S., Elizabeth A. Newnham, Dina Birman, Robert Lee, B. Heidi Ellis, and Christopher M. Layne. 2017. "Comparing Trauma Exposure, Mental Health Needs, and Service Utilization across Clinical Samples of Refugee, Immigrant, and U.S.-Origin Children." *Journal of Traumatic Stress* 30 (3): 209-18. <https://doi.org/10.1002/jts.22186>.
- Bong, Mimi, and Einar M. Skaalvik. 2003. "Academic Self-Concept and Self-Efficacy: How Different Are They Really?" *Educational Psychology Review* 15 (1): 1-40. <https://doi.org/10.1023/A:1021302408382>.
- Brophy, Marcia. 2017. *An Unbearable Reality: The Impact of War and Displacement on Children's Mental Health in Iraq*. London: Save the Children. [https://resourcecentre.savethechildren.net/node/12182/pdf/iraq\\_an\\_unbearable\\_reality\\_june\\_2017\\_1.pdf](https://resourcecentre.savethechildren.net/node/12182/pdf/iraq_an_unbearable_reality_june_2017_1.pdf).
- Burde, Dana, Amy Kapit, Rachel L. Wahl, Ozen Guven, and Margot Iglan Skarpeteig. 2017. "Education in Emergencies: A Review of Theory and Research." *Review of Educational Research* 87 (3): 619-58. <https://doi.org/10.3102/0034654316671594>.
- CASEL. 2015. *What Is Social and Emotional Learning?* Chicago: CASEL.
- Chirwa, Grames, and Devika Naidoo. 2014. "Structural and Social Constraints in the Teaching of Life Skills for HIV/AIDS Prevention in Malawi Primary Schools." *South African Journal of Childhood Education* 4 (1): 61-79. <https://doi.org/10.4102/sajce.v4i1.67>.

- Darling-Churchill, Kristen E., and Laura Lippman. 2016. "Early Childhood Social and Emotional Development: Advancing the Field of Measurement." *Journal of Applied Developmental Psychology* 45:1-7. <https://doi.org/10.1016/j.appdev.2016.02.002>.
- Denham, Susanne A., Peter Ji, and Bridget Hamre. 2010. *Compendium of Preschool through Elementary School Social-Emotional Learning and Associated Assessment Measures*. Chicago: CASEL. <https://casel.org/wp-content/uploads/2016/06/compendium-of-preschool-through-elementary-school-social-emotional-learning-and-associated-assessment-measures.pdf>.
- Denham, Susanne A., Erin Way, Sara C. Kalb, Heather K. Warren-Khot, and Hideko H. Bassett. 2013. "Preschoolers' Social Information Processing and Early School Success: The Challenging Situations Task." *British Journal of Developmental Psychology* 31 (2): 180-97. <https://doi.org/10.1111/j.2044-835X.2012.02085.x>.
- Dodge, Kenneth A., Patrick S. Malone, Jennifer E. Lansford, Emma Sorbring, Ann T. Skinner, Sombat Tapanya, Liliana Maria Uribe Tirado et al. 2015. "Hostile Attributional Bias and Aggressive Behavior in Global Context." *Proceedings of the National Academy of Sciences* 112 (30): 9310-15. <https://doi.org/10.1073/pnas.1418572112>.
- D'Sa, Nikhit. 2019a. *International Social & Emotional Learning Assessment: Adaptation and Administration Guidance*. Washington, DC: Save the Children. [https://inee.org/resources/international-social-and-emotional-learning-assessment-isela?webform\\_id=toolkit\\_resources](https://inee.org/resources/international-social-and-emotional-learning-assessment-isela?webform_id=toolkit_resources).
- D'Sa, Nikhit. 2019b. "Adopt, Contextualize, or Adapt? Understanding the Complexities of Modifying or Developing a Measure of Children's Social and Emotional Competencies." In *NISSEM Global Briefs: Education for the Social, the Sustainable and the Emotional*, edited by Aaron Benavot, Jean Bernard, Colette Chabbott, S. Garnett Russell, Margaret Sinclair, Andy Smart, and James H. Williams, 60-71. Washington, DC: NISSEM.
- Durable Solutions Platform. 2019. *Far from Home: Future Prospects for Syrian Refugees in Iraq*. Durable Solutions Platform. <https://reliefweb.int/report/iraq/far-home-future-prospects-syrian-refugees-iraq>.

- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions." *Child Development* 82 (1): 405-32. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>.
- Ferrás, Silvia Diazgranados, Jamie Weiss-Yagoda, and Carly Tubbs Dolan. 2019. "Measuring Social-Emotional Learning Outcomes for Children in Crisis: An Education in Emergencies Open-Source Measurement Library." In *NISSEM Global Briefs: Education for the Social, the Sustainable and the Emotional*, edited by Aaron Benavot, Jean Bernard, Colette Chabbott, S. Garnett Russell, Margaret Sinclair, Andy Smart, and James H. Williams, 752-59. Washington, DC: NISSEM.
- Goodman, Anna, and Robert Goodman. 2009. "Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health." *Journal of the American Academy of Child and Adolescent Psychiatry* 48 (4): 400-03. <https://doi.org/10.1097/CHI.0b013e3181985068>.
- Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry* 38 (5): 581-86. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>.
- Greenberg, Mark T., Elene E. Domitrovich, Roger P. Weissberg, and Joseph A. Durlak. 2017. "Social and Emotional Learning as a Public Health Approach to Education." *The Future of Children* 27 (1): 13-32. <https://doi.org/10.1353/foc.2017.0001>.
- Haggerty, Kevin, Jenna Elgin, Andrew Woolley. 2010. *Social-Emotional Learning Assessment Measures for Middle School Youth*. Seattle: Raikes Foundation. <https://www.casel.org/wp-content/uploads/2016/01/DAP-Raikes-Foundation-Review-1.pdf>.
- Halle, Tamara G., and Kristen E. Darling-Churchill. 2016. "Review of Measures of Social and Emotional Development." *Journal of Applied Developmental Psychology* 45:8-18. <https://doi.org/10.1016/j.appdev.2016.02.003>.
- Halman, Paul G. 2019. "Challenges in Gathering Data after Natural Disasters and the Potential for Local Knowledge: The Case of the Nepal Earthquake 2015." *NORRAG Special Issue 02*, 28-30.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2-3): 61-83. <https://doi.org/10.1017/S0140525X0999152X>.

INEE (Inter-agency Network for Education in Emergencies) and EASEL (Ecological Approaches to Social Emotional Learning) Laboratory. 2020. *SEL and PSS Measurement and Assessment Tools in Education in Emergencies: Identifying, Analyzing, and Mapping Tools to Global Guidance Documents*. New York: INEE. <https://inee.org/resources/sel-and-pss-measurement-and-assessment-tools-education-emergencies-identifying-analyzing>.

Jones, Damon E., Mark Greenberg, and Max Crowley. 2015. "Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness." *American Journal of Public Health* (November): e1-8. <https://doi.org/10.2105/AJPH.2015.302630>.

Lange-Nielsen, Ida Ingridsdatter, Silje Koltveit, Abdel Aziz Mousa Thabet, Atle Dyregrov, Ståle Pallesen, Tom Backer Johnsen, and Jon Christian Laberg. 2012. "Short-Term Effects of a Writing Intervention among Adolescents in Gaza." *Journal of Loss and Trauma* 17 (5): 403-22. <https://doi.org/10.1080/15325024.2011.650128>.

Lemerise, Elizabeth A., and William F. Arsenio. 2000. "An Integrated Model of Emotion Processes and Cognition in Social Information Processing." *Child Development* 71 (1): 107-18. <https://doi.org/10.1111/1467-8624.00124>.

MacCallum, Robert C., Michael W. Browne, and Hazuki M. Sugawara. 1996. "Power Analysis and Determination of Sample Size for Covariance Structure Modeling." *Psychological Methods* 1 (2): 130-49. <https://doi.org/10.1037/1082-989X.1.2.130>.

Markus, Hazel, and Paula Nurius. 1986. "Possible Selves." *American Psychologist* 41 (9): 954-69. <https://doi.org/10.1037/0003-066X.41.9.954>.

McDonald, Alun, Misty Buswell, Sonia Khush, and Marcia Brophy. 2017. *Invisible Wounds: The Impact of Six Years of War on the Mental Health of Syria's Children*. London: Save the Children. [https://resourcecentre.savethechildren.net/node/10384/pdf/invisible\\_wounds.pdf](https://resourcecentre.savethechildren.net/node/10384/pdf/invisible_wounds.pdf).

McHugh, Mary L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22 (3): 276-82. <https://doi.org/10.11613/BM.2012.031>.

- McNatt, Zahirah, Neil Boothby, Mike Wessells, and Rita Lo. 2018. *Guidance on PSS Facilitating Psychosocial Wellbeing and Social and Emotional Learning*. New York: INEE. <https://inee.org/resources/inee-guidance-note-psychosocial-support>.
- Mellor, David, and Kathleen A. Moore. 2014. "The Use of Likert Scales with Children." *Journal of Pediatric Psychology* 39 (3): 369-79. <https://doi.org/10.1093/jpepsy/jst079>.
- Miller, Kenneth E., and Andrew Rasmussen. 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 and Medicine* 70 (1): 7-16. <https://doi.org/10.1016/j.socscimed.2009.09.029>.
- Ministry of Education and Sports. 2010. *Life Skills Curriculum for Primary School Teachers in Uganda*. Kampala: Government of Uganda. <https://healtheducationresources.unesco.org/library/documents/life-skills-curriculum-primary-school-teachers-uganda>.
- Morgos, Dorothy J., William Worden, and Leila Gupta. 2007. "Psychosocial Effects of War Experiences among Displaced Children in Southern Darfur." *Omega: Journal of Death and Dying* 56 (3): 229-53. <https://doi.org/10.2190/OM.56.3.b>.
- Oyserman, Daphna, and Stephanie Fryberg. 2006. "The Possible Selves of Diverse Adolescents: Content and Function across Gender, Race and National Origin." In *Possible Selves: Theory, Research, and Applications*, edited by Curtis Dunkel and Jennifer Kerpelman, 17-39. Hauppauge, NY: Nova Science Publishers.
- Payton, John, Roger P. Weissberg, Joseph A. Durlak, Allison B. Dymnicki, Rebecca D. Taylor, Kriston B. Schellinger, and Molly Pachan. 2008. *The Positive Impact of Social and Emotional Learning for Kindergarten to Eighth-Grade Students: Findings from Three Scientific Reviews*. Chicago: CASEL. <https://casel.org/wp-content/uploads/2016/06/the-positive-impact-of-social-and-emotional-learning-for-kindergarten-to-eighth-grade-students-technical-report.pdf>.
- Pisani, Lauren, Ivelina Borisova, and Amy Jo Dowd. 2018. "Developing and Validating the International Development and Early Learning Assessment (IDELA)." *International Journal of Educational Research* 91 (January): 1-15. <https://doi.org/10.1016/j.ijer.2018.06.007>.
- Preacher, Kristopher J., and Donna L. Coffman. 2006. "Computing Power and Minimum Sample Size for RMSEA." <http://quantpsy.org/rmsear/rmsear.htm>.

- Puerta, Maria Laura Sánchez, Alexandria Valerio, and Marcela Gutiérrez Bernal. 2016. *Taking Stock of Programs to Develop Socioemotional Skills: A Systematic Review of Program Evidence*. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-0872-2>.
- Qouta, Samir R., Esa Palosaari, Marwan Diab, and Raija-Leena Punamaki. 2012. "Intervention Effectiveness among War-Affected Children: A Cluster Randomized Controlled Trial on Improving Mental Health." *Journal of Traumatic Stress* 25:188-298. <https://doi.org/10.1002/jts.21707>.
- Ring, Hannah Reeves, and Amy R. West. 2015. "Teacher Retention in Refugee and Emergency Settings: The State of the Literature." *International Education Journal: Comparative Perspectives* 14 (3): 106-21.
- Royeen, Charlotte Basic. 1985. "Adaptation of Likert Scaling for Use with Children." *Occupational Therapy Journal of Research* 5 (1): 59-69. <https://doi.org/10.1177/153944928500500104>.
- Rutter, Michael. 1981. "Stress, Coping and Development: Some Issues and Some Questions." *Journal of Child Psychology and Psychiatry* 22 (4): 323-56. <https://doi.org/10.1111/j.1469-7610.1981.tb00560.x>.
- Schreiber, James B., Frances K. Stage, Jamie King, Amaury Nora, and Elizabeth A. Barlow. 2006. "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review." *The Journal of Educational Research* 99 (6): 323-37. <https://doi.org/10.3200/JOER.99.6.323-338>.
- Schultz, David, Carroll E. Izard, and George Bear. 2004. "Children's Emotion Processing: Relations to Emotionality and Aggression." *Development and Psychopathology* 16:371-87. <https://doi.org/10.1017/S0954579404044566>.
- Selman, Robert L. 2003. *The Promotion of Social Awareness: Powerful Lessons from the Partnership of Developmental Theory and Classroom Practice*. New York: Russell Sage Foundation.
- Sklad, Marcin, Rene Diekstra, Monique De Ritter, and Jehonathan Ben. 2012. "Effectiveness of School-Based Universal Social, Emotional, and Behavioral Programs: Do They Enhance Students' Development in the Area of Skill, Behavior, and Adjustment?" *Psychology in the Schools* 49 (9): 892-909. <https://doi.org/10.1002/pits.21641>.

- Smith, Patrick, Sean Perrin, William Yule, Berima Hacam, and Rune Stuvland. 2002. "War Exposure among Children from Bosnia-Herzegovina: Psychological Adjustment in a Community Sample." *Journal of Traumatic Stress* 15 (2): 147-56. <https://doi.org/10.1023/A:1014812209051>.
- Stewart, Jay. 2013. "Tobit or Not Tobit?" *Journal of Economic and Social Measurement* 38: 263-90. <https://doi.org/10.3233/JEM-130376>.
- Streiner, David L. 2003. "Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency." *Journal of Personality Assessment* 80 (1): 99-103. [https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18).
- Super, Charles M., and Sara Harkness. 1986. "The Developmental Niche: A Conceptualization at the Interface of Child and Culture." *International Journal of Behavioral Development* 9 (4): 545-69. <https://doi.org/10.1177/016502548600900409>.
- Taylor, Rebecca D., Eva Oberle, Joseph A. Durlak, and Roger P. Weissberg. 2017. "Promoting Positive Youth Development through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-up Effects." *Child Development* 88 (4): 1-16. <https://doi.org/10.1111/cdev.12864>.
- Torrente, Catalina, John Lawrence Aber, Leighann Starkey, Brian Johnston, Anjali Shivshanker, Nina Weisenhorn, Jeannie Annan, Edward Seidman, Sharon Wolf, and Carly Tubbs Dolan. 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): 413-47. <https://doi.org/10.1080/19345747.2018.1561963>.
- Ungar, Michael, Gerry Connelly, Linda Liebenberg, and Linda Theron. 2017. "How Schools Enhance the Development of Young People's Resilience." *Social Indicators Research* 145: 615-27. <https://doi.org/10.1007/s11205-017-1728-8>.
- Upadhyay, Ashish Kumar, Swati Srivastava, and Chhavi Paul. 2017. "Consequences of Forced Migration during Early Childhood on Cognitive Well-Being in Later Childhood in Andhra Pradesh, India." *International Journal of Population Studies* 3 (2): 16. <https://doi.org/10.18063/ijps.v3i2.348>.

- Von Culin, Katherine R., Eli Tsukayama, and Angela L. Duckworth. 2014. "Unpacking Grit: Motivational Correlates of Perseverance and Passion for Long-Term Goals." *The Journal of Positive Psychology* 9 (4): 306-12. <https://doi.org/10.1080/17439760.2014.898320>.
- Wang, Huan, James Chu, Prashant Loyalka, Tao Xin, Yaojiang Shi, Qinghe Qu, and Chu Yang. 2016. "Can Social-Emotional Learning Reduce School Dropout in Developing Countries?" *Journal of Policy Analysis and Management* 35 (4): 818-47. <https://doi.org/10.1002/pam.21915>.
- Winthrop, Rebecca, and Jackie Kirk. 2008. "Learning for a Bright Future: Schooling, Armed Conflict, and Children's Well-Being." *Comparative Education Review* 52 (4): 639-61. <https://doi.org/10.1086/591301>.
- Woerner, Wolfgang, Bacy Fleitlich-Bilyk, Rhonda Martinussen, Janet Fletcher, Giulietta Cucchiario, Paulo Dalgalarrodo, Mariko Lui, and Rosemary Tannock. 2004. "The Strengths and Difficulties Questionnaire Overseas: Evaluations and Applications of the SDQ beyond Europe." *European Child and Adolescent Psychiatry*, Supplement 13 (2): 47-54. <https://doi.org/10.1007/s00787-004-2008-0>.
- Zinsser, Katherine M., Roger P. Weissberg, and Linda Dusenbury. 2013. *Aligning Preschool through High School Social and Emotional Learning Standards: A Critical and Doable Next Step*. Chicago: CASEL. <https://files.eric.ed.gov/fulltext/ED574866.pdf>.