Psychometric Evidence on a Measure of School-Related Stress and Stress Reactivity among Conflict-Affected Nigerian Refugee and Nigerien Local Children in Southern Niger: Response to Stress Questionnaire (RSQ)[1]

Ha Yeon Kim[2], Zezhen Wu[1]
NYU Global TIES for Children
Technical Working Paper
June 2022

The program implementation and impact evaluation of Healing Classrooms in Niger was funded by Dubai Cares. We thank the E-Cubed Research Envelope for supporting the time of the authors to work on this paper. The Evidence for Education in Emergencies (E-Cubed) Research Fund aims to strengthen the evidence base in Education in Emergencies, by supporting contextually relevant and usable research, and disseminating global public goods. Dubai Cares partnered with Inter-agency Network for Education in Emergencies (INEE) in 2017 to design and manage this research fund. To learn more, please visit https://inee.org/evidence/e-cubed. We also thank the New York University Abu Dhabi Research Institute for core funding to the New York University Global TIES for Children center, which provides support for staff and activities that directly contributed to the research presented here. In addition, we thank Dubai Cares and Economic and Social Research Council (ESRC) and Department for International Development (DFID) for supporting continued analyses of the datasets described in this manuscript.

[1] This paper is available at https://inee.org/measurement-library/response-stress-questionnaire-rsq-niger.

[2] Correspondence should be directed to Ha Yeon Kim at hayeon@nyu.edu
Global TIES for Children, New York University
627 Broadway, Room 807, New York, NY, USA 10012 | Steinhardt.nyu.edu/ihdsc/global-ties

Abstract

This study aims to provide evidence on the psychometric properties of the Response to Stress Questionnaire (RSQ) – Academic Problems scale, tested with refugee and local children in Niger. We tested two subscales (Academic Problems Stress, Involuntary Engagement) of the RSQ that were originally developed by Connor-Smith et al. (2000) for use with U.S. adult and child populations experiencing various stress situations and adapted for use with refugee and local children in Niger. Data were collected from a sample of 897 children (aged 5-16), randomly selected from the 1,795 second to fourth grade children who participated in a large cluster-randomized trial of the International Rescue Committee's (IRC) remedial and social and emotional learning (SEL) programming in Diffa, Niger. Psychometric analyses, including factor analysis, correlational analysis, and measurement invariance tests, indicated that the measure captures two unique dimensions of children's response to stress: *School-Related Stress Experience* and *Stress Reactivity*. We provide evidence that the two subscales of RSQ–Academic Problems tested in this study measure these dimensions with good reliability. We also provide evidence that the measure functions and is understood in the same way by children: with and without access to SEL programming; at the beginning, middle, and the end of the school year; and across gender. In addition, the two dimensions captured in this questionnaire were correlated in the expected directions within each wave, providing additional evidence of validity. However, while scores on each dimension were positively correlated across the three time points within a school year, the magnitude of the correlations was low – suggesting that children's school-related stress experience and stress reactivity might vary across the school year.

Introduction

In the context of the UN Sustainable Development Goals (SDGs), national and global stakeholders are working to ensure equitable access to high-quality education for refugee children that leads to relevant and effective learning outcomes (The World Bank, 2016; United Nations, 2016). To ensure the sustainability of such efforts, since 2012, stakeholders have sought to include refugees in national education systems in host countries (UNHCR, 2012). Evidence suggests, however, that the schooling experience of children affected by conflict and crisis, for both refugee and local children, can be often fraught, especially in a context like the Diffa region of Niger, where the host community is also affected by conflict and crisis.

Since the mid-2010s, violent attacks by Boko Haram—a jihadist militant group whose name translates to "Western education is forbidden"—resulted in approximately 213,000 Nigerian refugees and internally-displaced Nigeriens seeking protection in Niger's Diffa region. The violence and atrocity targeting modern education resulted in severe disruptions to Nigerian and Nigerien children's education and schooling experiences. Among vulnerable refugee and internally displaced persons (IDP) populations in Diffa, nearly 50% are school-aged children under the age of 18. Though public (formal) schools are open to all children, Nigerian refugee children do not have the necessary French (which is the official instructional language of Niger) language skills to follow instruction, and many local and IDP children also experienced interruption in their schooling due to persistent displacement and the evolving security situation as well as resource scarcity.

These stressful academic experiences may contribute: in the short-term, to high drop-out and low attendance rates; and in the long-term, to disruptions in the development of children's adaptive stress coping skills as well as other social, emotional, and academic and physical skills (Laurent et al., 2015; Nygaard et al., 2012; Sheridan et al., 2017; Trickett et al., 2011). Therefore, it is important for the practitioners working with refugee children to be able to assess and understand children's academic stress experiences in their daily lives.

However, we are aware of few field-feasible measures of children's academic stress response that have been adapted and tested for use with refugee children in low- and middle-income host- country contexts. In this paper, we provide psychometric evidence of a measure adapted to capture refugee children's stress experiences and potentially maladaptive stress responses in reaction to common stressors they experience in schools: the Response to Stress Questionnaire (RSQ)–Academic Problems. There exists evidence of this measure being tested on a large sample of Syrian refugee children enrolled in public schools in Lebanon (Kim, Wu, Gjicali, & Tubbs Dolan, 2020). In this study, we test the Response to Stress Questionnaire with a sample of both refugee and local children aged 5-16 attending Nigerien public schools. Before turning to the psychometric evidence on the measure, we first review the types of stress responses and how the RSQ was originally developed to measure those responses.

Types of Stress Responses

In the face of multiple stressors in their lives, children may exhibit various involuntary responses to stress. Involuntary responses refer to the emotional and physiological reactions towards stress that are difficult to intentionally control, even when there is conscious awareness of the reactions (Compas, 2000). For example, when facing academic setbacks in school, a child may involuntarily become frustrated and thus anxious with the problem (i.e., emotional reactions) or develop physical symptoms such as increased blood pressure (i.e., physiological reactions). In contrast, voluntary responses refer to conscious efforts to cope with stress and self-regulate accordingly (Compas, 2000). For example, when facing the same setbacks, the child can also actively seek solutions or help to cope with the stress.

Despite limited evidence on how children exactly develop ways to respond to and cope with stress, children do develop typical stress coping skills over time (Wadsworth, 2015). In their early years, children may engage in many "temperamentally based" involuntary stress reactions such as crying and involuntary self-soothing behavior (Compas et al., 2001; Connor-Smith et al., 2000) to regulate their physiology, emotion, and behavior. These reactions are automatic and arguably innate (Compas et al., 2001; Compas et al., 2004). Then, as children become more emotionally and cognitively self-aware through interactions with the people and environment in preschool and primary school, they continue to develop throughout their childhood and early adolescence various advanced cognitive strategies (such as cognitive reappraisal) and behavioral strategies (such as the intentional deployment of coping skills) (Zimmer-Gembeck & Skinner, 2011). By adolescence, children are expected to have developed a set of coping strategies that they can match onto the stressors in context (Wadsworth, 2015), even though they may still exhibit involuntary stress responses from time to time (Epstein-Ngo et al., 2013). Thus, through exposure to mild stress over time, children should be able to gradually develop more voluntary strategies to cope with stressful events, rather than simply resorting to automatic and involuntary responses.

But what could happen if this normative developmental cycle is interrupted? The traumatic and stressful events that refugee children can experience pre-, peri, and post-migration may elicit involuntary responses rather than volitional coping strategies (Zimmer-Gembeck & Skinner, 2011). In turn, children may have difficulty developing effective coping strategies given that meaningful daily interactions with mild stress may be overshadowed by traumatic experiences. As a result, they may tend to resort to more automatic responses to stress (Ullman & Peter- Hagene, 2014), such as negative emotional and physiological responses or complete avoidance and emotional numbing. At the same time, involuntary responses to stress are also associated with higher degrees of internalizing and externalizing symptoms (Wadsworth et al., 2005), which can subsequently exacerbate the perceived stress. Therefore, it is crucial for researchers to identify and measure the level of involuntary responses among refugee children to understand this important construct in their adaptation into host-country schools.

Measure Description

The Response to Stress Questionnaire (RSQ: Connor-Smith et al., 2000) was developed in an attempt to capture the ways that individuals react to and cope with specific sources of stress, including parental depression, family conflict, and academic stressors. It originally consisted of two parts: (a) *stress experience,* or the level of stress experienced about a specific target domain; and (b) *response to stress*, or how often the respondent uses or experiences each type of stress response in reaction to the specific stress experiences reported in the first part of the measure. The response to stress section of the original measure consists of 57 items corresponding to three subscales assessing voluntary coping strategies (primary control coping, secondary control coping, and disengagement coping) and two subscales assessing involuntary responses (involuntary engagement and involuntary disengagement). Previous studies using RSQ reported moderate to high test-retest reliability and internal consistency of the response to stress section of the measure across multiple samples (test-retest reliability = 0.69-0.81 across subscales; Cronbach's alpha = 0.67-0.92 across subscales and different stressors; Connor-Smith et al., 2000). Evidence of internal consistency was also established on various samples of the adolescent population (e.g., Navajo adolescents, alpha > 0.79; Wadsworth et al., 2004) or on the same subscales as the original study but adapted for a different stressor (e.g., economic stress and parental conflict, alpha > 0.67; Compas et al., 2001).

In order to capture refugee children's stress experiences and stress responses in public school settings for program evaluation purposes, we adapted the child self-report version of the RSQ-Academic Problems. We specifically included 10 items on students' school-related stress experiences (labeled as RSQS) and 15 items on their involuntary emotional and physiological responses and engagement to school-related stressors (labeled as RSIE). This subscale was chosen because it captures a type of stress response linked to adverse outcomes and likely to be displayed by younger children and in extreme stress situations. The same scales were also used and evaluated with Syrian refugee children in Lebanon, and have shown good psychometric properties acceptable for evaluation purposes (see Kim et al., 2021).

Table 1: Item description

RSQS	
1 = Not At All; 2 = A Little; 3 = Somewhat; 4 = Very	
RSQS1	Doing badly on a test or paper
RSQS2	Getting bad grades or report cards
RSQS3	Not understanding classes

RSQS4	Not understanding homework
RSQS5	Feeling pressured to do something
RSQS6	Having bad classes or teachers
RSQS7	Having trouble studying
RSQS8	Not having your homework done
RSQS9	Teachers that yell or get angry
RSQS10	Pressure from parents or teachers to perform perfectly

RSIE

Below is a list of things that children and teenagers sometimes do, think, or feel when they are dealing with school problems. Everyone deals with problems in their own way – some people do a lot of the things on this list or have a bunch of feelings, other people just do or think a few of these things.
Think of all the stressful parts of school that you indicated above. For each item below, circle one number from 1 (not at all) to 4 (a lot) that shows how much you do or feel these things when you have problems with school like the ones you indicated above. Please let us know about everything you do, think, and feel, even if you don't think it helps make things better.
The items also are rated on a 4-point scale (1 = not at all, 2 = a little, 3 = somewhat, 4 = a lot)

Physical arousal: 1, 3, 6
Intrusive thoughts: 2, 4, 5
Rumination: 7, 10, 13
Emotional arousal: 9, 11, 14
Involuntary action: 8, 12, 15

RSIE1	1. When dealing with school problems, I feel sick to my stomach or get headaches
RSIE2	2. I keep remembering the school problems that happened or can't stop thinking about what might happen
RSIE3	3. I get really jumpy when I am dealing with the stress of school problems
RSIE4	4. When I am trying to sleep, I can't stop thinking about the stressful aspect of school problems that happened or I have bad dreams about them
RSIE5	5. Thoughts about school problems just pop up into my head

RSIE6	6. When I’m dealing with school problems, I feel it in my body: (a) My heart races (b) I feel hot or sweaty (c) My breathing speeds up (d) My muscles get tight
RSIE7	7. When something stressful happens related to school problems, I can’t stop thinking about how I am feeling
RSIE8	8. When something stressful happens related to school problems, I can’t always control what I do
RSIE9	9. When I am faced with school problems, right away I feel really: (a) Angry (b) Sad (c) Worried/anxious (d) Scared
RSIE10	10. After something stressful happens related to school problems, I can’t stop thinking about what I did or said
RSIE11	11. When something stressful happens related to school problems, I get upset by things that don’t usually bother me
RSIE12	12. When stressful things happen related to school problems I sometimes act without thinking
RSIE13	13. When something stressful happens related to school problems, I can’t stop thinking about why this is happening
RSIE14	14. My thoughts start racing when I am faced with school problems
RSIE15	15. When I am dealing with the stress of school problems, sometimes I can’t control what I do or say

Data and Sample

The current data were collected on students from 30 public schools from a list of 75 potential sites located in the towns of Diffa and Maine-Soroa in Niger. The following criteria were utilized for school selection: (1) security clearance (2) distance from NGO Office <40km (3) sufficient numbers of teachers (more than eight teachers per school) and students (more than 120 students enrolled per school) (4) serves a minimum of three primary grades (some schools did not serve the full range of elementary grade levels).

Of the 30 schools selected, 20 schools were traditional French-only schools and the other 10 schools were French-Arabic schools; French-Arabic schools often had a religious focus, with instructional emphasis on Arabic in order to read the Koran, in addition to French. Eighteen schools were located in Diffa, and 12 schools were located in Maine-Soroa. The student composition of the schools varied widely, reflecting the ongoing refugee crisis and ethnic/linguistic diversity of the Diffa region. Specifically, schools ranged from 10-42% of refugee or internally displaced students, 0-85% Kanuri speakers, 0-48% Hausa speakers, and 2-95% Fulfulde speakers, with smaller percentages of the student body speaking other home languages. The majority of second to fourth grade students in our sample struggled academically, with 75-100% of students unable to read Grade one level texts in French and 61-98% of students unable to solve simple subtraction problems in the screening tests.

The student participants IIe 897 children randomly selected from the 1,795 second to fourth grade children who participated in a large cluster-randomized trial of the IRC's remedial and SEL programming. To minimize students' assessment burden while capturing a wide range of children's academic and social emotional functioning, some measures, including RSQ (Package A), were assessed only on a randomly-selected half of the sample (see Table 2 for descriptive statistics on the student participants). All participants of the study (N = 1795) were enrolled in the participating schools and offered the IRC's Healing Classrooms Remedial Tutoring programming (53% girls, 19% refugees). The participants were selected through a screening test and lottery. First, all second to fourth graders (N=5,684) were assessed on their French literacy and mathematics skills using an adapted version of the assessment used for the Annual Status of Education Report (ASER: Banerji, Bhattacharjea, & Wadhwa, 2013). Of them, 4,994 children (96% of second graders, 85% of third graders, and 79% of fourth graders) were deemed eligible for tutoring based on their low performance in both French literacy (unable to read Grade One level texts in French) and mathematics (unable to solve simple subtraction problems) test. Secondly, due to the large number of students who met the eligibility criteria, a total of 1,800 students (90 classrooms across 30 schools, 20 students per classroom) were randomly selected to enroll in the tutoring program. In consultation with the local staff, we selected an equal number of eligible students from each grade level, when possible. This was for ease of classroom grouping and to give (somewhat) of a higher chance for children in older grades to receive tutoring (i.e., those in more urgent need for tutoring support, given grade and proficiency levels) while maintaining a perception of fairness and avoiding stigmatization. Students who met tutoring eligibility requirements but were not selected via lottery (n = 3,194) were excluded from the study. Of the randomly selected 1,800 children, 5 children were found to be either duplicate entries (i.e., listed twice on the student list) or unidentifiable due to administrative errors, and therefore excluded from the sample.

Table 2: Sample descriptives (n = 897)

	Mean (S.D.) / %	Min	Max
Age	9.18 (1.40)	6	15
Female	52.51%	0	1
Refugee	18.95%	0	1
Kanuri speakers	39.25%	0	1
Hausa speakers	84.06%	0	1
Fulfulde speakers	52.51%	0	1
Other language speakers	21.74%	0	1

Method

For the purpose of program evaluation, child assessments were conducted via verbal interviews to increase engagement and address literacy challenges. Assessments were recorded using tablets. Translations to French were made by local IRC staff in Niger and data were collected at 3 time points (November 2016, March 2017, and May 2017) by assessors who were trained by the researchers and local IRC staff to provide additional clarifications and explanations of specific concepts and vocabulary that may not be familiar to younger children.

Enumerators were recruited from an existing IRC database of data surveyors used in past projects, as well as from Zinder University. All applicants were required to take French and Hausa literacy screening tests. The highest-scoring applicants were offered a position in the training. Due to security restrictions on expats in the Diffa region, enumerators were trained in Niamey by NYU-TIES researchers and IRC local staff (See the accompanied RSQ training materials). Training activities included an introduction to the measure and a description of it by NYU-TIES and IRC staff using PowerPoint slides and a measure manual, followed by staff demonstrations, then volunteer enumerators' group practice, in which they practiced administering the RSQ. Measures were introduced and practiced first in French, which is the official language of Niger, and a common language among enumerators. Next, standardized translations into local Diffa languages of Hausa and Kanuri were agreed upon with enumerators and IRC staff with the whole-group. Official Hausa and Kanuri translators were hired to attend midline training to further facilitate this process. After all training of the research measures were completed, enumerators pilot practiced in the field with local children in Diffa.

All descriptive, bivariate correlation and reliability analyses were conducted using Stata SE version 15.1, and all measurement modeling was conducted using Mplus 8.3 (Muthén & Muthén, 2018). In order to account for the structural characteristics of the data, two important specifications were made for all measurement models.

First, given the 4-point scale item response options in the measure, items were specified as categorical. Because modeling categorical responses as normally and continuously distributed can lead to an inflation of model fit statistics and biased estimation of factor loadings and standard errors, we used a weighted least squares mean and variance-adjusted (WLSMV) estimator with a probit-link function (Beauducel & Herzberg, 2006; Lei, 2007).

Second, we used robust standard errors to adjust for clustering because 1) students were nested within classrooms/teachers, and classrooms/teachers within sites; and 2) it was an effective and efficient way to model complex data when sample size at the cluster level was not small (Huang, 2016). In all models, model fits were evaluated using Hu & Bentler's (1999) criteria: RMSEA (Root Mean Square Error Of Approximation) < 0.06, CFI (Comparative Fit Index) < 0.95, TLI (Tucker–Lewis Index) < 0.95, SRMR (Standardized Root Mean Squared Residual) < 0.08. Missing data were pairwise deleted (i.e., all available information was used from all cases) to preserve the full sample (Asparouhov & Muthén, 2010). As a result, we were able to include and obtain factor scores for all children who were assessed for any items of RSQ in the analysis regardless of missing information on specific items.

Results

Descriptive Statistics

RSQS items are generally negatively skewed at all waves, with 1 (*not at all*) being less frequently chosen as compared to the other values indicative of higher school-related stress.

RISE items at all waves are generally positively skewed with few students having “a lot of” stress reactivity. RSIE6 and RSIE9 are the exceptions because they each have four options where the students need to rate from 1 to 4, and the maximum value among the ratings of the four options is taken to represent the intensity of physical responses and negative emotions in general. Therefore, there are more ratings of 3s and 4s than 1s and 2s for these two items at all waves.

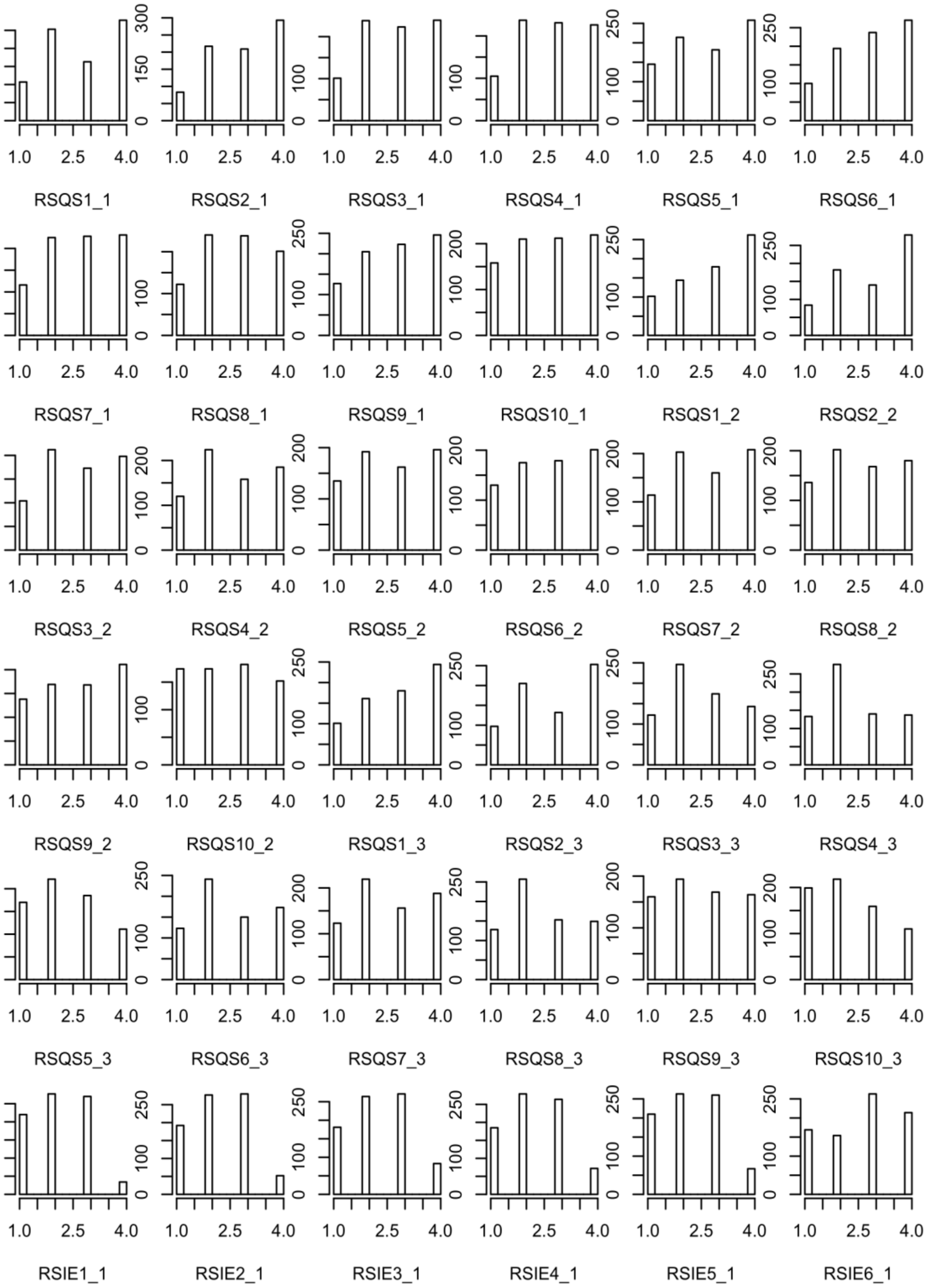
Table 3: Descriptive statistics

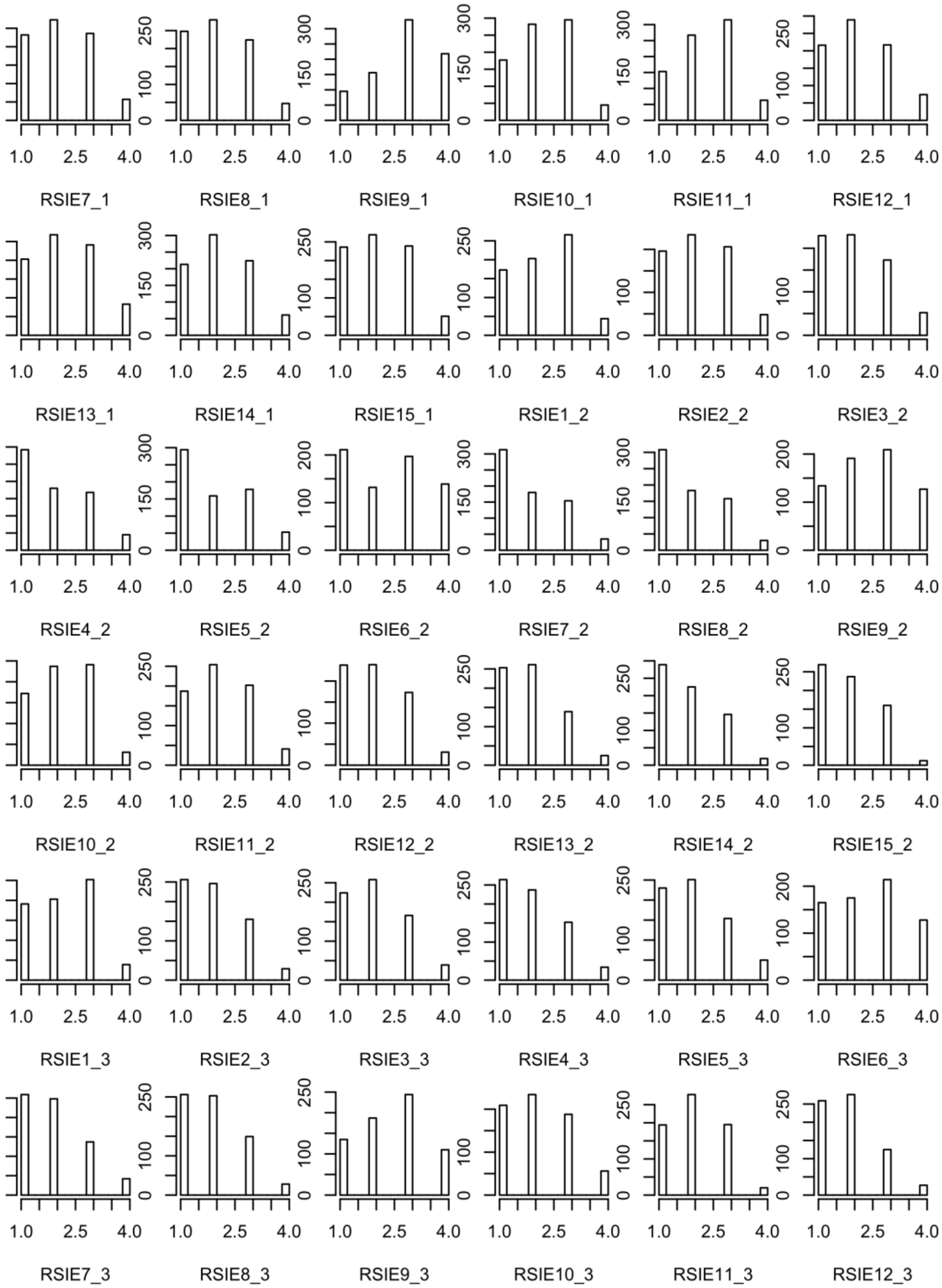
variable	n_missing	complete_rate	mean	sd	*p0*	*p25*	*p50*	*p75*	*p100*
RSQS1_1	96	0.893	2.764	1.069	1	2	3	4	4
RSQS2_1	95	0.894	2.888	1.019	1	2	3	4	4
RSQS3_1	96	0.893	2.749	1.019	1	2	3	4	4
RSQS4_1	98	0.891	2.723	1.015	1	2	3	4	4
RSQS5_1	98	0.891	2.692	1.106	1	2	3	4	4
RSQS6_1	96	0.893	2.845	1.028	1	2	3	4	4
RSQS7_1	97	0.892	2.717	1.035	1	2	3	4	4
RSQS8_1	96	0.893	2.647	1.018	1	2	3	4	4
RSQS9_1	96	0.893	2.734	1.062	1	2	3	4	4
RSQS10_1	98	0.891	2.616	1.087	1	2	3	4	4
RSQS1_2	210	0.766	2.875	1.082	1	2	3	4	4
RSQS2_2	212	0.764	2.896	1.074	1	2	3	4	4
RSQS3_2	210	0.766	2.677	1.049	1	2	3	4	4
RSQS4_2	210	0.766	2.594	1.063	1	2	2	4	4
RSQS5_2	212	0.764	2.612	1.098	1	2	3	4	4
RSQS6_2	212	0.764	2.658	1.092	1	2	3	4	4
RSQS7_2	212	0.764	2.674	1.078	1	2	3	4	4

variable	n_missing	complete_rate	mean	sd	*p0*	*p25*	*p50*	*p75*	*p100*
RSQS8_2	211	0.765	2.571	1.081	1	2	3	4	4
RSQS9_2	211	0.765	2.659	1.115	1	2	3	4	4
RSQS10_2	215	0.760	2.457	1.098	1	1	2	3	4
RSQS1_3	211	0.765	2.827	1.073	1	2	3	4	4
RSQS2_3	210	0.766	2.787	1.090	1	2	3	4	4
RSQS3_3	212	0.764	2.493	1.013	1	2	2	3	4
RSQS4_3	211	0.765	2.410	1.015	1	2	2	3	4
RSQS5_3	210	0.766	2.345	1.023	1	2	2	3	4
RSQS6_3	210	0.766	2.543	1.054	1	2	2	4	4
RSQS7_3	211	0.765	2.596	1.072	1	2	3	4	4
RSQS8_3	210	0.766	2.470	1.028	1	2	2	3	4
RSQS9_3	210	0.766	2.491	1.093	1	2	2	3	4
RSQS10_3	211	0.765	2.262	1.047	1	1	2	3	4
RSIE1_1	96	0.893	2.147	0.872	1	1	2	3	4
RSIE2_1	96	0.893	2.240	0.890	1	2	2	3	4
RSIE3_1	98	0.891	2.320	0.938	1	2	2	3	4
RSIE4_1	97	0.892	2.279	0.919	1	2	2	3	4
RSIE5_1	97	0.892	2.230	0.933	1	1	2	3	4
RSIE6_1	97	0.892	2.652	1.089	1	2	3	4	4
RSIE7_1	99	0.890	2.148	0.923	1	1	2	3	4
RSIE8_1	98	0.891	2.088	0.905	1	1	2	3	4
RSIE9_1	99	0.890	2.840	0.959	1	2	3	4	4
RSIE10_1	98	0.891	2.260	0.866	1	2	2	3	4
RSIE11_1	99	0.890	2.361	0.879	1	2	2	3	4
RSIE12_1	101	0.887	2.187	0.939	1	1	2	3	4
RSIE13_1	102	0.886	2.257	0.955	1	1	2	3	4
RSIE14_1	97	0.892	2.166	0.908	1	1	2	3	4

variable	n_missing	complete_rate	mean	sd	*p0*	*p25*	*p50*	*p75*	*p100*
RSIE15_1	102	0.886	2.132	0.915	1	1	2	3	4
RSIE1_2	211	0.765	2.264	0.910	1	1	2	3	4
RSIE2_2	213	0.763	2.155	0.920	1	1	2	3	4
RSIE3_2	212	0.764	2.070	0.942	1	1	2	3	4
RSIE4_2	212	0.764	1.950	0.966	1	1	2	3	4
RSIE5_2	213	0.763	1.985	1.001	1	1	2	3	4
RSIE6_2	218	0.757	2.389	1.127	1	1	2	3	4
RSIE7_2	215	0.760	1.870	0.935	1	1	2	3	4
RSIE8_2	218	0.757	1.867	0.920	1	1	2	3	4
RSIE9_2	236	0.737	2.498	1.020	1	2	3	3	4
RSIE10_2	216	0.759	2.192	0.868	1	1	2	3	4
RSIE11_2	213	0.763	2.142	0.889	1	1	2	3	4
RSIE12_2	216	0.759	1.996	0.887	1	1	2	3	4
RSIE13_2	219	0.756	1.906	0.847	1	1	2	2	4
RSIE14_2	218	0.757	1.845	0.854	1	1	2	2	4
RSIE15_2	219	0.756	1.875	0.830	1	1	2	3	4
RSIE1_3	212	0.764	2.203	0.914	1	1	2	3	4
RSIE2_3	211	0.765	1.937	0.875	1	1	2	3	4
RSIE3_3	210	0.766	2.029	0.892	1	1	2	3	4
RSIE4_3	210	0.766	1.936	0.895	1	1	2	3	4
RSIE5_3	212	0.764	2.035	0.923	1	1	2	3	4
RSIE6_3	215	0.760	2.447	1.053	1	2	3	3	4
RSIE7_3	211	0.765	1.945	0.906	1	1	2	3	4
RSIE8_3	211	0.765	1.926	0.866	1	1	2	3	4
RSIE9_3	221	0.754	2.487	0.988	1	2	3	3	4
RSIE10_3	210	0.766	2.132	0.942	1	1	2	3	4
RSIE11_3	210	0.766	2.060	0.825	1	1	2	3	4

variable	n_missing	complete_rate	mean	sd	*p0*	*p25*	*p50*	*p75*	*p100*
RSIE12_3	210	0.766	1.884	0.839	1	1	2	2	4
RSIE13_3	213	0.763	1.940	0.880	1	1	2	3	4
RSIE14_3	212	0.764	1.909	0.910	1	1	2	3	4
RSIE15_3	213	0.763	1.895	0.863	1	1	2	3	4


RSQS1_1
RSQS2_1
RSQS3_1
RSQS4_1
RSQS5_1
RSQS6_1
RSQS7_1
RSQS8_1
RSQS9_1
RSQS10_1
RSQS1_2
RSQS2_2
RSQS3_2
RSQS4_2
RSQS5_2
RSQS6_2
RSQS7_2
RSQS8_2
RSQS9_2
RSQS10_2
RSQS1_3
RSQS2_3
RSQS3_3
RSQS4_3
RSQS5_3
RSQS6_3
RSQS7_3
RSQS8_3
RSQS9_3
RSQS10_3
RSIE1_1
RSIE2_1
RSIE3_1
RSIE4_1
RSIE5_1
RSIE6_1


RSIE7_1
RSIE8_1
RSIE9_1
RSIE10_1
RSIE11_1
RSIE12_1
RSIE13_1
RSIE14_1
RSIE15_1
RSIE1_2
RSIE2_2
RSIE3_2
RSIE4_2
RSIE5_2
RSIE6_2
RSIE7_2
RSIE8_2
RSIE9_2
RSIE10_2
RSIE11_2
RSIE12_2
RSIE13_2
RSIE14_2
RSIE15_2
RSIE1_3
RSIE2_3
RSIE3_3
RSIE4_3
RSIE5_3
RSIE6_3
RSIE7_3
RSIE8_3
RSIE9_3
RSIE10_3
RSIE11_3
RSIE12_3

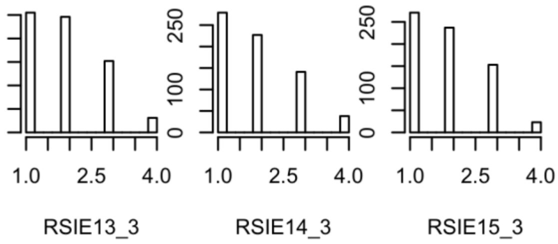


Figure 1: Item distribution

Factor Analysis

EFA and CFA

Before beginning analyses, we randomly divided our sample in half in order to create exploratory and confirmatory samples at each time point. Exploratory samples were used to examine multiple versions of data-driven models, of which a final proposed solution was selected based on conceptual and empirical considerations. Confirmatory samples were used to test the proposed factor structure, thereby building confidence in the stability of empirically derived exploratory factor analytic estimates (Osborne & Fitzpatrick, 2012). CFA models with a good model fit and the same factor structure across baseline, midline, and endline were used as final models for subsequent analysis.

First, we performed EFA models to empirically explore the factor structure. The scree plot of the eigenvalues obtained during the EFA step suggested a 2-or 3-factor structure (see Figure 2). Upon close examination of the items, two RSIE items were deemed either contextually and culturally inappropriate, likely to be confounded with other factors, or difficult to respond for children in Diffa, Niger, based on field reports: (a) RSIE1: When dealing with school problems, I feel sick to my stomach or get headaches (may be confounded with food insecurity and physical illness, which is common in the Niger context); and (b) RSIE9: When I am faced with school problems, right away I feel really: (a) Angry (b) Sad (c) Worried/anxious (d) Scared (different ethnic groups and cultures within a sample have different levels of acceptance of emotional expression; some languages spoken in Diffa do not distinguish and/or do not have the words for certain emotions; and young children were unable to identify different emotions).

After removing these two items, CFA models confirmed that the originally hypothesized 2-factor structure – including one factor representing students' school-related stress experiences (RSQS) and a second factor representing stress reactivity (involuntary engagement) responses to stress (RSIE) – had a good fit across all waves of the data. Given the consistency with the existing measure and good model fit across waves, this two-factor model was used for the subsequent analysis.

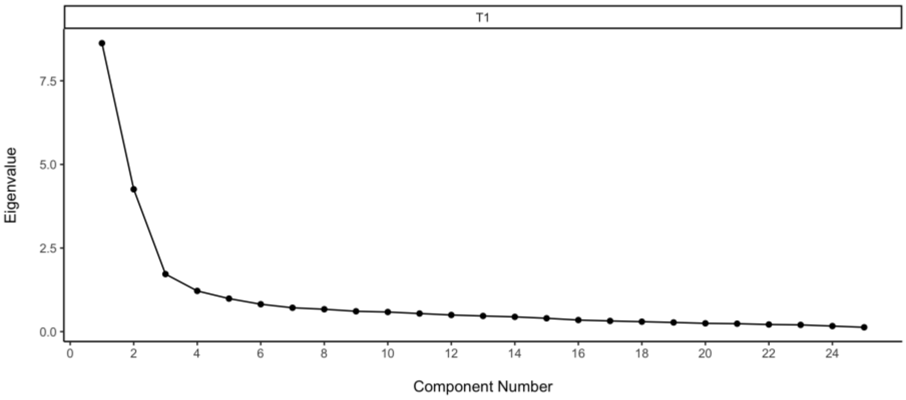


Figure 2: EFA model screeplots at all waves

Table 4: CFA model fits at all waves

k	χ^2	df	*p*	CFI	TLI	RMSEA	SRMR	Filename
93	606.467	229	0	0.948	0.943	0.062	0.069	RSQ1_CFA2-ie19.out
93	474.784	229	0	0.971	0.968	0.054	0.056	RSQ2_CFA2-ie19.out
93	670.116	229	0	0.951	0.946	0.073	0.070	RSQ3_CFA2-ie19.out

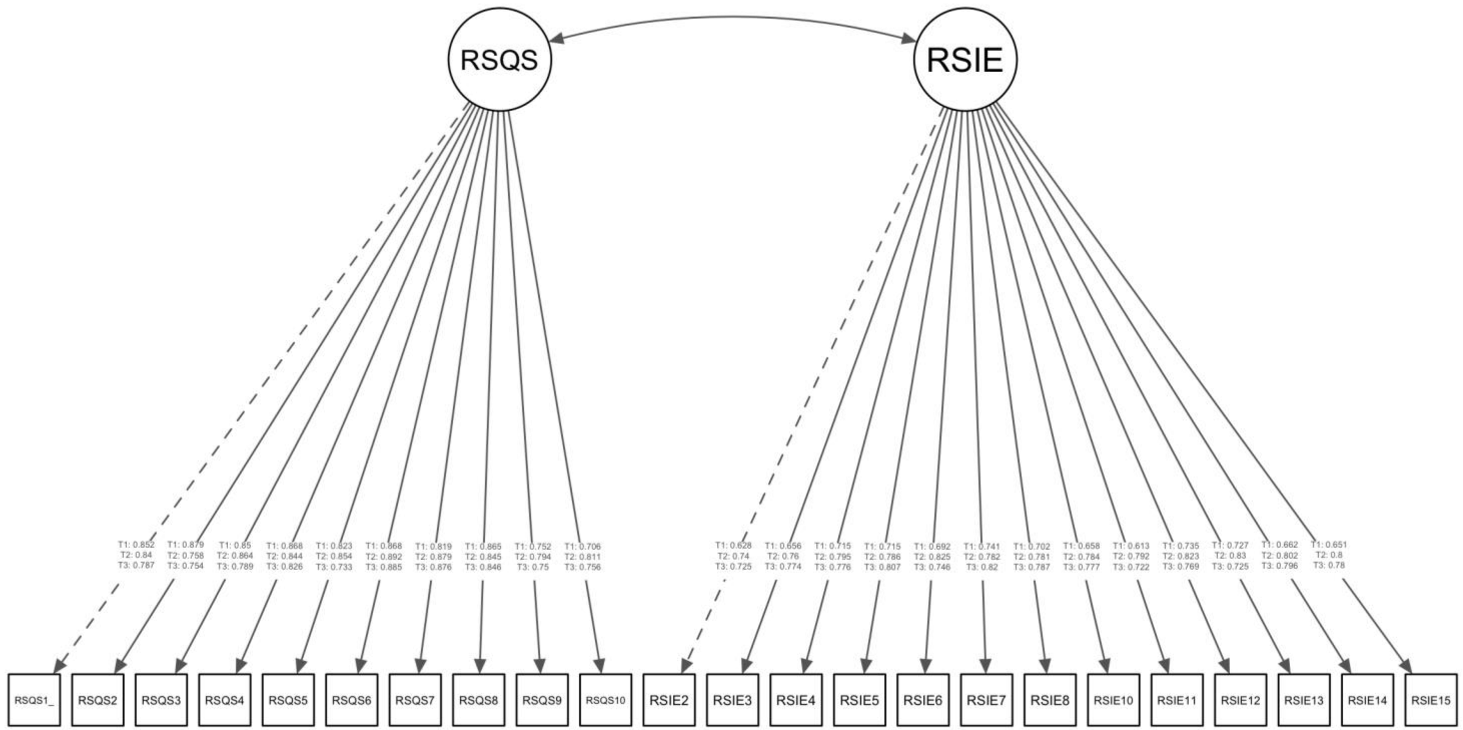


Figure 3: CFA model final factor structure

Table 5: CFA model parameters at all waves

paramHeader	param	est_T1	se_T1	est_se_T1	pval_T1	est_T2	se_T2	est_se_T2	pval_T2	est_T3	se_T3	est_se_T3	pval_T3
RSQS.BY	RSQS1	0.852	0.016	54.522	0	0.840	0.024	35.233	0	0.787	0.026	29.793	0
RSQS.BY	RSQS2	0.879	0.014	62.642	0	0.758	0.033	22.744	0	0.754	0.030	25.345	0
RSQS.BY	RSQS3	0.850	0.017	49.093	0	0.864	0.021	40.745	0	0.789	0.023	33.639	0
RSQS.BY	RSQS4	0.868	0.017	49.959	0	0.844	0.023	36.017	0	0.826	0.022	37.678	0
RSQS.BY	RSQS5	0.823	0.022	37.946	0	0.854	0.023	37.198	0	0.733	0.028	25.909	0
RSQS.BY	RSQS6	0.868	0.014	60.302	0	0.892	0.017	52.774	0	0.885	0.018	49.562	0
RSQS.BY	RSQS7	0.819	0.023	35.107	0	0.879	0.020	43.448	0	0.876	0.019	45.956	0
RSQS.BY	RSQS8	0.865	0.017	50.316	0	0.845	0.023	37.175	0	0.846	0.023	37.089	0
RSQS.BY	RSQS9	0.752	0.025	30.101	0	0.794	0.026	30.927	0	0.750	0.030	25.249	0
RSQS.BY	RSQS10	0.706	0.030	23.583	0	0.811	0.025	33.002	0	0.756	0.027	28.474	0
RSIE.BY	RSIE2	0.628	0.034	18.613	0	0.740	0.029	25.362	0	0.725	0.027	26.479	0
RSIE.BY	RSIE3	0.656	0.037	17.804	0	0.760	0.027	28.534	0	0.774	0.025	31.473	0
RSIE.BY	RSIE4	0.715	0.025	28.327	0	0.795	0.029	27.654	0	0.776	0.024	32.799	0
RSIE.BY	RSIE5	0.715	0.030	23.532	0	0.786	0.026	29.831	0	0.807	0.021	38.708	0
RSIE.BY	RSIE6	0.692	0.040	17.290	0	0.825	0.026	31.558	0	0.746	0.029	25.647	0
RSIE.BY	RSIE7	0.741	0.026	28.303	0	0.782	0.027	28.504	0	0.820	0.021	38.978	0
RSIE.BY	RSIE8	0.702	0.026	27.203	0	0.781	0.028	27.844	0	0.787	0.021	37.150	0

paramHeader	param	est_T1	se_T1	est_se_T1	pval_T1	est_T2	se_T2	est_se_T2	pval_T2	est_T3	se_T3	est_se_T3	pval_T3
RSIE.BY	RSIE10	0.658	0.035	18.980	0	0.784	0.027	28.793	0	0.777	0.023	33.889	0
RSIE.BY	RSIE11	0.613	0.036	17.208	0	0.792	0.022	35.924	0	0.722	0.028	25.492	0
RSIE.BY	RSIE12	0.735	0.029	25.559	0	0.823	0.022	37.904	0	0.769	0.024	32.215	0
RSIE.BY	RSIE13	0.727	0.027	26.944	0	0.830	0.019	43.480	0	0.725	0.025	28.978	0
RSIE.BY	RSIE14	0.662	0.027	24.739	0	0.802	0.023	34.259	0	0.796	0.021	37.697	0
RSIE.BY	RSIE15	0.651	0.032	20.299	0	0.800	0.027	30.162	0	0.780	0.023	33.464	0
RSIE.WITH	RSQS	0.229	0.059	3.872	0	0.310	0.047	6.545	0	0.469	0.040	11.810	0
Variances	RSQS	1.000	0.000	999.000	999	1.000	0.000	999.000	999	1.000	0.000	999.000	999
Variances	RSIE	1.000	0.000	999.000	999	1.000	0.000	999.000	999	1.000	0.000	999.000	999

Table 6: CFA model R-squared at all waves

param	est_T1	se_T1	est_se_T1	pval_T1	scale_f_T1	est_T2	se_T2	est_se_T2	pval_T2	scale_f_T2	est_T3	se_T3	est_se_T3	pval_T3	scale_f_T3
RSQS1	0.727	0.027	27.261	0	0.523	0.705	0.040	17.616	0	0.543	0.619	0.042	14.896	0	0.617
RSQS2	0.772	0.025	31.321	0	0.478	0.575	0.051	11.372	0	0.652	0.568	0.045	12.673	0	0.657
RSQS3	0.722	0.029	24.547	0	0.527	0.746	0.037	20.372	0	0.504	0.623	0.037	16.819	0	0.614
RSQS4	0.753	0.030	24.980	0	0.497	0.713	0.040	18.009	0	0.536	0.682	0.036	18.839	0	0.564
RSQS5	0.677	0.036	18.973	0	0.568	0.730	0.039	18.599	0	0.520	0.537	0.041	12.955	0	0.680
RSQS6	0.754	0.025	30.151	0	0.496	0.795	0.030	26.387	0	0.453	0.783	0.032	24.781	0	0.466
RSQS7	0.671	0.038	17.553	0	0.574	0.773	0.036	21.724	0	0.477	0.767	0.033	22.978	0	0.483
RSQS8	0.748	0.030	25.158	0	0.502	0.714	0.038	18.587	0	0.534	0.715	0.039	18.545	0	0.534
RSQS9	0.565	0.038	15.050	0	0.660	0.630	0.041	15.463	0	0.608	0.562	0.045	12.625	0	0.662
RSQS10	0.499	0.042	11.792	0	0.708	0.658	0.040	16.501	0	0.585	0.572	0.040	14.237	0	0.654
RSIE2	0.394	0.042	9.307	0	0.779	0.548	0.043	12.681	0	0.673	0.525	0.040	13.239	0	0.689
RSIE3	0.430	0.048	8.902	0	0.755	0.578	0.041	14.267	0	0.650	0.599	0.038	15.737	0	0.633
RSIE4	0.512	0.036	14.163	0	0.699	0.633	0.046	13.827	0	0.606	0.602	0.037	16.400	0	0.631
RSIE5	0.511	0.043	11.766	0	0.699	0.618	0.041	14.916	0	0.618	0.651	0.034	19.354	0	0.590
RSIE6	0.479	0.055	8.645	0	0.722	0.681	0.043	15.779	0	0.565	0.557	0.043	12.824	0	0.666
RSIE7	0.549	0.039	14.152	0	0.671	0.611	0.043	14.252	0	0.624	0.672	0.034	19.489	0	0.573
RSIE8	0.494	0.036	13.601	0	0.712	0.610	0.044	13.922	0	0.625	0.619	0.033	18.575	0	0.617

param	est_T1	se_T1	est_se_T1	pval_T1	scale_f_T1	est_T2	se_T2	est_se_T2	pval_T2	scale_f_T2	est_T3	se_T3	est_se_T3	pval_T3	scale_f_T3
RSIE10	0.432	0.046	9.490	0	0.753	0.614	0.043	14.397	0	0.621	0.603	0.036	16.945	0	0.630
RSIE11	0.376	0.044	8.604	0	0.790	0.628	0.035	17.962	0	0.610	0.521	0.041	12.746	0	0.692
RSIE12	0.541	0.042	12.779	0	0.678	0.677	0.036	18.952	0	0.569	0.591	0.037	16.108	0	0.639
RSIE13	0.528	0.039	13.472	0	0.687	0.690	0.032	21.740	0	0.557	0.526	0.036	14.489	0	0.688
RSIE14	0.439	0.035	12.369	0	0.749	0.643	0.038	17.129	0	0.597	0.634	0.034	18.849	0	0.605
RSIE15	0.424	0.042	10.150	0	0.759	0.640	0.042	15.081	0	0.600	0.608	0.036	16.732	0	0.626

Internal Reliability and Correlations

To assess internal consistency, Cronbach's alpha (α) of each latent factor within each data collection time point was calculated. We also assessed McDonald's omega (ω; Hayes & Coutts, 2020; McDonald, 1999) of each latent factor as a more general reliability estimate that does not assume equal factor loadings (i.e., tau-equivalence). The recommendation from the contemporary literature of assessing reliability for unidimensional measures assuming unequal factor loadings, like RSQ, is to avoid α and use ω (Revelle & Zinbarg, 2009; Zinbarg et al., 2005). Therefore, we report both reliability statistics but focus on the interpretation of ω. While there are no definitive and universal guidelines for interpreting α and ω, α > 0.7 is generally accepted as acceptable/high reliability, and Nájera Catalán (2019) suggests a higher standard for ω > 0.8 as excellent evidence of internal consistency.

Table 7 presents both the unweighted (Cronbach's α) and the weighted (McDonald's ω) internal consistency estimates of the RSQS and RSIE scales. Overall, both subscales have high internal consistency at all waves for both Cronbach's α and McDonald's ω statistics.

Table 7: Item total statistics

subscale	raw_alpha	std_alpha	G6(smc)	average_r	S/N	alpha_se	mean	sd	median_r	omega_lg	omega_by_wave
wave: T1											
RSQS	0.931	0.931	0.931	0.576	13.569	0.003	2.737	0.820	0.600	0.954	0.957
wave: T2											
RSQS	0.939	0.939	0.939	0.606	15.395	0.003	2.669	0.870	0.609	0.961	0.960
wave: T3											
RSQS	0.923	0.923	0.926	0.545	11.977	0.004	2.523	0.807	0.549	0.950	0.947
wave: T1											
RSIE	0.885	0.885	0.896	0.340	7.723	0.006	2.288	0.575	0.330	0.909	0.920
wave: T2											
RSIE	0.927	0.929	0.935	0.465	13.014	0.004	2.067	0.659	0.463	0.951	0.956
wave: T3											
RSIE	0.928	0.929	0.934	0.464	12.994	0.004	2.051	0.640	0.467	0.949	0.950

Note. See below for the explanation of each statistics presented in the table above.
Raw_alpha : Cronbach's alpha (Cronbach, 1951) based upon the covariances
std_alpha : The standardized alpha based upon the correlations
G6(smc) : Guttman's Lambda 6 reliability
average_r : The average interitem correlation
S/N : Signal/Noise ratio
alpha_se : Standard error of alpha
var_r : The variance of the interitem correlations

median_r : The median of the interitem correlations
mean : The mean of the scale formed by averaging the items
sd : The standard deviation of the total score
omega_lg : McDonald's Omega (McDonald, 1999) with factor loadings and residuals extracted from the scalar longitudinal invariance model
omega_by_wave: McDonald's Omega (McDonald, 1999) with factor loadings and residuals extracted from each wave of CFA model

Measurement Invariance

We conducted (1) measurement invariance tests across treatment, gender, and refugee groups in each wave; and (2) longitudinal invariance testing across baseline, midline, and endline. Measurement invariance refers to the extent to which a set of items measures an underlying construct of interest in the same way across groups or times (Reise et al., 1993). If a measure operates or is understood differently in different groups, then one should not compare group differences on observed scores (Glanville & Wildhagen, 2007). For example, without evidence of measurement invariance, one should not compare boys' and girls' stress reactivity; compare this construct with and without access to SEL interventions; or track changes in students' stress reactivity over time.

For each set of analyses, we tested for levels of measurement invariance by fitting a series of nested models in which we progressively constrained the model parameters to equality across groups/time points. Specifically, we fit models within each time point and then across time points to test the equality of 1) the factor structure in treatment and control groups and time points (configural invariance); 2) the factor loadings across groups/time points (metric invariance); and 3) the item intercepts or thresholds across groups/time points (scalar invariance) (Gregorich, 2006; Millsap, 2011). We assessed the relative fit of each of these models against the configural model using criteria suggested by (Chen, 2007; metric invariance: ΔCFI <0.01; ΔRMSEA < 0.015 ΔSRMR < 0.030; scalar invariance: ΔCFI < 0.01, ΔRMSEA < 0.015, ΔSRMR < 0.010). If the imposition of equality constraints did not provide a significant decrement of model fit, we concluded that the hypothesis of invariance was supported.

Treatment invariance. We found evidence of scalar invariance at all waves between treatment and control groups (see Table 8 for model fits). This means that the latent factors across two different treatment groups measure equivalent constructs, and therefore we can directly compare treatment and control group students on the same RSQ scale, without bias.

Gender measurement invariance. We found that RSQ was scalar invariant at all waves across gender groups (see Table 9 for model fits), suggesting that we can compare mean differences by gender on the RSQ constructs without bias due to child gender.

Refugee measurement invariance. We found that RSQ was scalar invariant at baseline and midline across refugee status (see Tables 10 for model fits), suggesting that we can compare mean differences by refugee status on the RSQ constructs without bias due to their status.

Invariance across time. As shown in Table 11, a series of longitudinal invariance models were tested to confirm that changes from baseline to midline, and midline to endline, of the same construct can be estimated. Model fit difference between configural, metric, and scalar models suggested the factor structure, loadings, and thresholds of the items were invariant across baseline, midline, and endline. In other words, we found no significant difference in the item and measure functioning across waves, and we can compare baseline, midline, and endline scores on these constructs as assessed using the RSQ.

Table 8: Treatment group invariance model fit

k	χ^2	df	*p*	χ^2_B	df	*p*	$\Delta\chi^2$	df	*p*	CFI	TLI	RMSEA	SRMR	Filename
wave: T1														
186	966.214	458	0	8,422.686	506	0				0.936	0.929	0.053	0.068	RSQ1_tx_inv_config.out
165	978.641	479	0	8,422.686	506	0	47.837	21	0.0007	0.937	0.933	0.051	0.070	RSQ1_tx_inv_metric.out
98	1,043.306	546	0	8,422.686	506	0	90.428	67	0.0299	0.937	0.942	0.048	0.071	RSQ1_tx_inv_scalar.out
wave: T2														
186	857.172	458	0	10,747.893	506	0				0.961	0.957	0.050	0.059	RSQ2_tx_inv_config.out
165	870.958	479	0	10,747.893	506	0	47.719	21	0.0008	0.962	0.960	0.049	0.062	RSQ2_tx_inv_metric.out
98	930.167	546	0	10,747.893	506	0	82.434	67	0.0969	0.962	0.965	0.045	0.063	RSQ2_tx_inv_scalar.out
wave: T3														
186	1,018.313	458	0	15,659.553	506	0				0.963	0.959	0.060	0.067	RSQ3_tx_inv_config.out
165	1,008.352	479	0	15,659.553	506	0	44.537	21	0.0020	0.965	0.963	0.057	0.069	RSQ3_tx_inv_metric.out
98	1,063.269	546	0	15,659.553	506	0	70.247	67	0.3694	0.966	0.968	0.053	0.069	RSQ3_tx_inv_scalar.out

Table 9: Gender invariance model fit

k	χ^2	df	*p*	χ^2_B	df	*p*	$\Delta\chi^2$	df	*p*	CFI	TLI	RMSEA	SRMR	Filename
wave: T1														
186	1,054.385	458	0	11,541.03	506	0				0.946	0.940	0.057	0.066	rsq1_gen_inv_config.out
165	1,046.674	479	0	11,541.03	506	0	34.880	21	0.0291	0.949	0.946	0.054	0.068	rsq1_gen_inv_metric.out
98	1,107.219	546	0	11,541.03	506	0	73.901	67	0.2630	0.949	0.953	0.051	0.068	rsq1_gen_inv_scalar.out
wave: T2														
186	951.337	458	0	14,809.52	506	0				0.966	0.962	0.056	0.058	rsq2_gen_inv_config.out
165	937.579	479	0	14,809.52	506	0	33.231	21	0.0437	0.968	0.966	0.053	0.059	rsq2_gen_inv_metric.out
98	989.632	546	0	14,809.52	506	0	66.621	67	0.4901	0.969	0.971	0.049	0.060	rsq2_gen_inv_scalar.out
wave: T3														
186	1,167.838	458	0	18,145.23	506	0				0.960	0.956	0.067	0.066	rsq3_gen_inv_config.out
165	1,128.735	479	0	18,145.23	506	0	35.048	21	0.0279	0.963	0.961	0.063	0.068	rsq3_gen_inv_metric.out
98	1,182.486	546	0	18,145.23	506	0	61.685	67	0.6605	0.964	0.967	0.058	0.068	rsq3_gen_inv_scalar.out

Table 10: Refugee invariance model fit

k	χ^2	df	*p*	χ^2_B	df	*p*	$\Delta\chi^2$	df	*p*	CFI	TLI	RMSEA	SRMR	Filename
wave: T1														
186	973.671	458	0	10,147.82	506	0				0.947	0.941	0.053	0.068	rsq1_ref_inv_config.out
165	976.767	479	0	10,147.82	506	0	46.949	21	0.0010	0.948	0.945	0.051	0.069	rsq1_ref_inv_metric.out
98	1,027.714	546	0	10,147.82	506	0	75.914	67	0.2132	0.950	0.954	0.047	0.070	rsq1_ref_inv_scalar.out
wave: T2														
186	924.639	458	0	12,291.49	506	0				0.960	0.956	0.054	0.060	rsq2_ref_inv_config.out
165	942.889	479	0	12,291.49	506	0	53.802	21	0.0001	0.961	0.958	0.053	0.062	rsq2_ref_inv_metric.out
98	988.848	546	0	12,291.49	506	0	73.252	67	0.2805	0.962	0.965	0.049	0.062	rsq2_ref_inv_scalar.out
wave: T3														
186	1,111.592	458	0	18,387.57	506	0				0.963	0.960	0.064	0.067	rsq3_ref_inv_config.out
165	1,023.162	479	0	18,387.57	506	0	21.903	21	0.4051	0.970	0.968	0.058	0.067	rsq3_ref_inv_metric.out
98	1,078.626	546	0	18,387.57	506	0	69.633	67	0.3890	0.970	0.972	0.053	0.068	rsq3_ref_inv_scalar.out

Table 11: Longitudinal invariance model fit

k	χ^2	df	*p*	χ^2_B	df	*p*	$\Delta\chi^2$	df	*p*	CFI	TLI	RMSEA	WRMR	Filename
291	2,979.050	2,262	0	24,365.24	2,346	0				0.967	0.966	0.020	1.218	RSQ123_CFA2-ie19_inv_config.out
249	3,019.132	2,304	0	24,365.24	2,346	0	82.224	42	0.0002	0.968	0.967	0.019	1.245	RSQ123_CFA2-ie19_inv_metric.out
115	3,199.406	2,438	0	24,365.24	2,346	0	329.302	134	0.0000	0.965	0.967	0.019	1.289	RSQ123_CFA2-ie19_inv_scalar.out

Evidence of Correlational Validity

As seen in Table 12, correlations between the two subscales at each time point were moderate (*r*s =0.28-0.48), positive, and statistically significant at all waves. This indicates that students who felt more stressed by school-related issues were also likely to have more involuntary and reactive responses. On the other hand, the correlations across waves for each subscale were low (*r*s < 0.15), with correlation between Time 1 and Time 3 not significantly correlated. This suggests that scores on each of the RSQ subscale constructs were relatively time-variant and subject to change.

Table 12: Partial correlations among latent variables across time

	1	2	3	4	5	6
1. RSQS_1	--	--	--	--	--	--
2. RSQS_2	0.131**	--	--	--	--	--
3. RSQS_3	0.058	0.130**	--	--	--	--
4. RSIE_1	0.279***	0.125**	0.032	--	--	--
5. RSIE_2	0.055	0.301***	0.083	0.098*	--	--
6. RSIE_3	0.141***	0.077	0.480***	0.027	0.115*	--

** p < 0.05. ** p < 0.01. *** p < 0.001.*

Notes: Correlation estimates are obtained from the longitudinal invariance models.

Table 13: Bivariate correlations among factor scores

	1	2	3	4	5	6
1. RSQS_1	--					
2. RSQS_2	0.158***	--				
3. RSQS_3	0.069*	0.159***	--			
4. RSIE_1	0.278***	0.151***	0.062	--		
5. RSIE_2	0.074*	0.326***	0.115***	0.128***	--	
6. RSIE_3	0.176***	0.098**	0.513***	0.039	0.147***	--

** $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.*

Notes: Factor scores are obtained from the longitudinal invariance models

Conclusion

The RSQ, was used in this study to evaluate the impact of access to non-formal, SEL-infused tutoring programming among Nigerien refugee and Nigerien local students. Evidence indicates that this version of the RSQ holds promise for use as a program evaluation measure, with evidence of validity based on internal structural and correlational patterns; evidence of reliability based on internal consistency estimates; and evidence of measurement invariance across treatment groups, gender, refugee status, and time.

First, measures used for program evaluation purposes must have strong evidence of validity, or evidence that scores on the measure can be interpreted as capturing the focal constructs – in this case, stress and stress reactivity. Factor analyses of RSQ suggested two factors consistent with existing theory and prior empirical evidence about the structure of the measure. Exploratory and confirmatory factor analytic models provided a good fit to the data, and all items loaded highly onto the factors at all waves. This provides strong evidence for the internal structure validity of the measure.

Second, data from program evaluation measures must be highly reliable, as measurement error can attenuate the ability to detect program impact (Raudenbush & Sadoff, 2008). We found both subscales of the RSQ had high internal consistency, indicating that children generally gave consistent ratings on items within these RSQ subscales (αs > 0.80 and ωs > 0.90).

Third, data from program evaluation measures should also provide evidence that the measure is understood and reported in the same way and on the same scale by children in different treatment and demographic groups, as well as overtime. This criterion is known as measurement invariance. Establishing the measurement invariance of an assessment used in a rigorous program impact evaluation enables us to confidently assess whether children's responses to stress are improving or declining over time – and whether such changes are the result of our SEL programming (Halpin et al., 2019; Halpin & Torrente, 2014). RSQ has strong evidence of:

- Treatment invariance, indicating that one can compare the average scores of the treatment and control groups without bias due to treatment conditions.
- Gender invariance, indicating that one can compare the average scores of boys and girls without bias due to gender.
- Refugee invariance, suggesting that scores from the derived RSQ scale can be used to capture meaningful group differences in response to stress and stress reactivity by refugee status. This piece of evidence indicates that the measure is not biased when deriving mean differences by refugee status.

- Longitudinal invariance, indicating that one can directly assess growth overtime on the constructs assessed by the empirically derived RSQ subscales.

However, similar to the results from the RSQ's testing in Lebanon (Kim, Wu, Gjicali, & Tubbs Dolan, 2020) the correlations between the same constructs over time – approximately in 3-month intervals across a school year – were low to moderate for all subscales. We are unaware of other studies that have examined the longitudinal stability of the level of stress and stress reactivity among children using self-report methods, and therefore we are not able to interpret these findings in the context of prior evidence[3]. We also found very few studies that examined children's stress experiences and stress response longitudinally. Studies with physiological measures of stress, e.g., salivary cortisol level, suggest that children who are exposed to adversity in their early lives are likely to have acute and hyperactive responses to stress when exposed to stressful situations (Laurent et al., 2015; Trickett et al., 2010). In addition, longitudinal studies on child-report measures of post-traumatic stress reaction show a significant change in stress reactions over time, with low to moderate levels of correlations across time (Aziz & Vostanis, 2000; Nygaard, Jensen, & Dyb, 2012). Given the existing evidence, we argue children's stress experiences and their stress responses are not stable traits. Rather, they are temporary reactions to the immediate situation, and they are likely to change in response to the presence and absence of stressors in, and given their own adaptation to, the school environment.

In addition, it is important to note that the measure tested in this study does not include all of the items in the original RSQ measure (57 items), and we do not provide evidence of the other RSQ subscales (primary control, secondary control, disengagement coping, involuntary disengagement). Similarly, this study only provides evidence on the ability of the RSQ to measure the level of stress experiences and stress reactivity about school-related problems. For the use of the RSQ with other stressors, further evidence is needed.

Lastly, we were unable to conduct a comprehensive pilot and cognitive interviews prior to the assessment, due to limited time and resources, as well as challenges of testing a measure in different languages. During the analysis process later, we found two items deemed inappropriate for the context and culture, as reported in this report earlier. While this study was able to identify

---

[3] The only study we were able to find that tested repeated measures of the RSQ was Connor-Smith et al. (2000). Authors assessed the test-retest reliability of the RSQ among college students within a 1- to 2-week interval, and they reported that involuntary engagement subscale scores were highly correlated between the two waves of data collection (r = 0.81; stress experiences subscale reliability was not reported). Given the much shorter time interval and the older population in the Connor-Smith et al. (2000) study, it is not generalizable to the current study.

measures with sufficient evidence of reliability and validity by removing those items, early identification and adaptation of problematic items could have strengthened the measure.

Recommendations for the Use of the Response to Stress Questionnaire (RSQ)

While the evidence provided in this study largely supports the use of RSQ for evaluation purposes with Nigerian refugee and Nigerien local children in conflict-affected Niger, we largely repeat here the recommendations we made for the RSQ tested with Syrian refugee children in public schools in Lebanon (Kim et al., 2020), with additional point (#4):

1. We recommend future studies to explore triangulation with measuring several aspects of children's response to stress through multiple sources of reporting and methods (e.g., child-report, physiological arousal (e.g., heart rate with stressful stimuli), parent-report, teacher-report) to further validate and better understand the stress experience and response in the target population.
2. Given the findings suggesting that stress and stress response is likely time-varying, we recommend the prompt to refer to a shorter time frame (e.g., two weeks) than the currently suggested 6-month period. Perceived stress and stress responses are likely to be immediate reactions to the stressors present at the time, and especially for young children, it may be challenging to recall the stress level and responses over a lengthy period of time.
3. Small adjustments can be made to the measure to potentially mitigate the effect of cognitive anchoring biases such as the recency effect, and, in turn, improve the stability of children's ratings of their stress response. Such mitigation strategies can include embedding explicit prompts in the introductory statement to ask respondents to think back to their experience of stress in school for a certain period of time (e.g., *"Think back to this [particular school-related stressor] that happened over the past two weeks. In general, I feel…"*).
4. Lastly, we highly recommend careful review and comprehensive cognitive interviews and a pilot of the items in multiple languages spoken among the target population, to increase cultural and contextual fit and ensure validity. This is especially important in a context where multiple languages are spoken and the target population is culturally and ethnically diverse.

Appendix

The explanations of the summary item statistics by each subscale are as follows (if table is printed):

raw.r : The correlation of each item with the total score, not corrected for item overlap

std.r : The correlation of each item with the total score (not corrected for item overlap) if the items were all standardized

r.cor : Item whole correlation corrected for item overlap and scale reliability

r.drop: Item whole correlation for this item against the scale without this item

mean : The mean of each item

sd : The standard deviation of each item

Table 14: Summary item statistics by each subscale

item	n	raw.r	std.r	r.cor	r.drop	mean	sd
wave: T1							
subscale: RSQS							
RSQS1_1	801	0.805	0.804	0.783	0.751	2.764	1.069
RSQS2_1	802	0.813	0.814	0.796	0.764	2.888	1.019
RSQS3_1	801	0.806	0.808	0.785	0.756	2.749	1.019
RSQS4_1	799	0.803	0.807	0.784	0.754	2.723	1.015
RSQS5_1	799	0.801	0.799	0.770	0.745	2.692	1.106
RSQS6_1	801	0.823	0.825	0.804	0.777	2.845	1.028
RSQS7_1	800	0.786	0.788	0.758	0.731	2.717	1.035
RSQS8_1	801	0.815	0.818	0.798	0.767	2.647	1.018
RSQS9_1	801	0.724	0.722	0.683	0.654	2.734	1.062
RSQS10_1	799	0.680	0.677	0.627	0.599	2.616	1.087
wave: T2							
RSQS1_2	687	0.793	0.793	0.768	0.739	2.875	1.082
RSQS2_2	685	0.749	0.750	0.715	0.686	2.896	1.074
RSQS3_2	687	0.815	0.817	0.797	0.768	2.677	1.049

item	n	raw.r	std.r	r.cor	r.drop	mean	sd
RSQS4_2	687	0.822	0.823	0.806	0.775	2.594	1.063
RSQS5_2	685	0.788	0.788	0.757	0.733	2.612	1.098
RSQS6_2	685	0.850	0.850	0.835	0.809	2.658	1.092
RSQS7_2	685	0.850	0.850	0.834	0.809	2.674	1.078
RSQS8_2	686	0.824	0.824	0.803	0.777	2.571	1.081
RSQS9_2	686	0.793	0.790	0.761	0.737	2.659	1.115
RSQS10_2	682	0.751	0.749	0.711	0.687	2.457	1.098
wave: T3							
RSQS1_3	686	0.785	0.784	0.764	0.726	2.827	1.073
RSQS2_3	687	0.731	0.728	0.698	0.658	2.787	1.090
RSQS3_3	685	0.786	0.789	0.763	0.730	2.493	1.013
RSQS4_3	686	0.804	0.807	0.788	0.752	2.410	1.015
RSQS5_3	687	0.734	0.737	0.695	0.667	2.345	1.023
RSQS6_3	687	0.844	0.844	0.831	0.800	2.543	1.054
RSQS7_3	686	0.820	0.819	0.799	0.769	2.596	1.072
RSQS8_3	687	0.794	0.795	0.771	0.738	2.470	1.028
RSQS9_3	687	0.701	0.697	0.649	0.620	2.491	1.093
RSQS10_3	686	0.684	0.684	0.637	0.604	2.262	1.047
wave: T1							
subscale: RSIE							
RSIE1_1	801	0.554	0.561	0.520	0.478	2.147	0.872
RSIE2_1	801	0.604	0.606	0.567	0.531	2.240	0.890
RSIE3_1	799	0.620	0.617	0.581	0.545	2.320	0.938
RSIE4_1	800	0.652	0.651	0.621	0.584	2.279	0.919
RSIE5_1	800	0.652	0.648	0.618	0.581	2.230	0.933
RSIE6_1	800	0.665	0.649	0.624	0.582	2.652	1.089
RSIE7_1	798	0.673	0.672	0.650	0.608	2.148	0.923

item	n	raw.r	std.r	r.cor	r.drop	mean	sd
RSIE8_1	799	0.642	0.640	0.612	0.572	2.088	0.905
RSIE9_1	798	0.515	0.506	0.450	0.424	2.840	0.959
RSIE10_1	799	0.617	0.628	0.599	0.551	2.260	0.866
RSIE11_1	798	0.599	0.603	0.565	0.528	2.361	0.879
RSIE12_1	796	0.642	0.644	0.617	0.572	2.187	0.939
RSIE13_1	795	0.655	0.656	0.628	0.586	2.257	0.955
RSIE14_1	800	0.603	0.606	0.572	0.530	2.166	0.908
RSIE15_1	795	0.602	0.607	0.574	0.529	2.132	0.915
wave: T2							
RSIE1_2	686	0.591	0.594	0.548	0.527	2.264	0.910
RSIE2_2	684	0.706	0.704	0.673	0.650	2.155	0.920
RSIE3_2	685	0.723	0.719	0.693	0.671	2.070	0.942
RSIE4_2	685	0.724	0.719	0.699	0.673	1.950	0.966
RSIE5_2	684	0.723	0.713	0.695	0.669	1.985	1.001
RSIE6_2	679	0.750	0.733	0.720	0.693	2.389	1.127
RSIE7_2	682	0.723	0.719	0.701	0.671	1.870	0.935
RSIE8_2	679	0.723	0.723	0.705	0.673	1.867	0.920
RSIE9_2	661	0.583	0.569	0.524	0.506	2.498	1.020
RSIE10_2	681	0.734	0.742	0.722	0.690	2.192	0.868
RSIE11_2	684	0.761	0.763	0.745	0.716	2.142	0.889
RSIE12_2	681	0.729	0.735	0.716	0.680	1.996	0.887
RSIE13_2	678	0.726	0.733	0.715	0.677	1.906	0.847
RSIE14_2	679	0.703	0.712	0.691	0.652	1.845	0.854
RSIE15_2	678	0.720	0.731	0.712	0.674	1.875	0.830
wave: T3							
RSIE1_3	685	0.689	0.687	0.659	0.634	2.203	0.914
RSIE2_3	686	0.685	0.688	0.659	0.632	1.937	0.875

item	n	raw.r	std.r	r.cor	r.drop	mean	sd
RSIE3_3	687	0.693	0.691	0.665	0.639	2.029	0.892
RSIE4_3	687	0.721	0.721	0.699	0.672	1.936	0.895
RSIE5_3	685	0.745	0.740	0.724	0.694	2.035	0.923
RSIE6_3	682	0.733	0.721	0.702	0.673	2.447	1.053
RSIE7_3	686	0.751	0.750	0.733	0.704	1.945	0.906
RSIE8_3	686	0.721	0.722	0.700	0.671	1.926	0.866
RSIE9_3	676	0.546	0.539	0.490	0.470	2.487	0.988
RSIE10_3	687	0.765	0.765	0.749	0.719	2.132	0.942
RSIE11_3	687	0.706	0.710	0.683	0.659	2.060	0.825
RSIE12_3	687	0.709	0.714	0.690	0.660	1.884	0.839
RSIE13_3	684	0.692	0.696	0.672	0.639	1.940	0.880
RSIE14_3	685	0.735	0.737	0.720	0.685	1.909	0.910
RSIE15_3	684	0.719	0.724	0.704	0.671	1.895	0.863

References

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, *2010*, 1–10.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Catalán, H. E. N. (2019). Reliability, population classification and weighting in multidimensional poverty measurement: A Monte Carlo study. *Social Indicators Research*, *142*(3), 887–910.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Compas, B. E. (2000). *Adolescent/Child's Self-Report Responses to Stress*. https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/2804/2019/04/14195649/RSQ-Academic-Stress-Child-Self-Report-2.pdf

Compas, B. E., Connor-Smith, J., & Jaser, S. S. (2004). Temperament, Stress Reactivity, and Coping: Implications for Depression in Childhood and Adolescence. *Journal of Clinical Child & Adolescent Psychology*, *33*(1), 21–31. https://doi.org/10.1207/S15374424JCCP3301_3

Compas, B. E., Connor-Smith, J., Saltzman, H., Thomsen, A., & Wadsworth, M. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin*, *127*(1), 87–127.

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.

Epstein-Ngo, Q., Maurizi, L. K., Bregman, A., & Ceballo, R. (2013). In response to community violence: Coping strategies and involuntary stress responses among Latino adolescents. *Cultural Diversity & Ethnic Minority Psychology*, *19*(1), 38–49. https://doi.org/10.1037/a0029753

Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement: Assessing dimensionality and measurement invariance across race and ethnicity. *Educational and Psychological Measurement*, *67*(6), 1019–1041. https://doi.org/10.1177/0013164406299126

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Halpin, P. F., & Torrente, C. (2014). Measuring Critical Education Processes and Outcomes: Illustration from a Cluster Randomized Trial in the Democratic Republic of the Congo. *Society for Research on Educational Effectiveness*.

Halpin, P. F., Wolf, S., Yoshikawa, H., Rojas, N., Kabay, S., Pisani, L., & Dowd, A. J. (2019). Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Developmental Psychology*, *55*(1), 23.

Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

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.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education*, *84*(1), 175–196.

Laurent, H. K., Gilliam, K. S., Wright, D. B., & Fisher, P. A. (2015). Child anxiety symptoms related to longitudinal cortisol trajectories and acute stress responses: Evidence of developmental stress sensitization. *Journal of Abnormal Psychology*, *124*(1), 68–79. https://doi.org/10.1037/abn0000009

Lei, P.-W. (2007). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, *43*(3), 495. https://doi.org/10.1007/s11135-007-9133-z

McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Earlbaum Associates. *Inc., Mahwah, NJ*, 142–145.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance* (pp. xii, 355). Routledge/Taylor & Francis Group.

Muthén, L. K., & Muthén, B. (2018). Mplus. *The Comprehensive Modelling Program for Applied Researchers: User's Guide*, *5*.

Nygaard, E., Jensen, T. K., & Dyb, G. (2012). Stability of posttraumatic stress reaction factors and their relation to general mental health problems in children: A longitudinal study. *Journal of Clinical Child & Adolescent Psychology*, *41*(1), 15–26. https://doi.org/10.1080/15374416.2012.632344

Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, *17*(15). https://doi.org/10.7275/h0bd-4d11

Raudenbush, S. W., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness*, *1*(2), 138–154.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145.

Sheridan, M. A., Peverill, M., Finn, A. S., & McLaughlin, K. A. (2017). Dimensions of childhood adversity have distinct associations with neural systems underlying executive functioning. *Development and Psychopathology*, *29*(5), 1777–1794. https://doi.org/10.1017/S0954579417001390

The World Bank. (2016, April). *Education and fragility* [Text/HTML]. World Bank. http://www.worldbank.org/en/topic/education/brief/education-and-fragility

Trickett, P. K., Negriff, S., Ji, J., & Peckins, M. (2011). Child maltreatment and adolescent development. *Journal of Research on Adolescence*, *21*(1), 3–20. https://doi.org/10.1111/j.1532-7795.2010.00711.x

Ullman, S. E., & Peter-Hagene, L. (2014). Social reactions to sexual assault disclosure, coping, perceived control, and PTSD symptoms in sexual assault victims. *Journal of Community Psychology*, *42*(4), 495–508. https://doi.org/10.1002/jcop.21624

United Nations. (2016). *The Sustainable Development Goals Report*. United Nations. http://www.un.org.lb/Library/Assets/The-Sustainable-Development-Goals-Report-2016-Global.pdf

United Nations High Commissioner for Refugees. (2012). *Education strategy 2012-2016*. UNHCR. https://www.unhcr.org/protection/operations/5149ba349/unhcr-education-strategy-2012-2016.html

Wadsworth, M. E., Rieckmann, T., Benson, M. A., & Compas, B. E. (2004). Coping and responses to stress in Navajo adolescents: Psychometric properties of the Responses to Stress Questionnaire. *Journal of Community Psychology*, *32*(4), 391–411. https://doi.org/10.1002/jcop.20008

Wadsworth, M. E., Raviv, T., Compas, B. E., & Connor-Smith, J. K. (2005). Parent and adolescent responses to poverty related stress: Tests of mediated and moderated coping models. *Journal of Child and Family Studies*, *14*(2), 283–298. https://doi.org/10.1007/s10826-005-5056-2

Wadsworth, M. E. (2015). Development of maladaptive coping: A functional adaptation to chronic, uncontrollable stress. *Child Development Perspectives*, *9*(2), 96–100. https://doi.org/10.1111/cdep.12112

Zimmer-Gembeck, M. J., & Skinner, E. A. (2011). Review: The development of coping across childhood and adolescence: An integrative review and critique of research. *International Journal of Behavioral Development*, *35*(1), 1–17. https://doi.org/10.1177/0165025410384923

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7