

Development of the Student Learning in Emergency Checklist (26)  
(SLEC-26): A measurement tool of promoters and barriers for learning  
among conflict-affected students in Palestine

June T. Forsberg<sup>1</sup>, Jon-Håkon Schultz<sup>1</sup>, Camilla Lodi<sup>2</sup>, & Carly Tubbs Dolan<sup>3</sup>

<sup>1</sup> Department of Education, University of Tromsø, The Arctic University of Norway

<sup>2</sup> Norwegian Refugee Council, Palestine

<sup>3</sup> NYU Global TIES for Children

Technical Working Paper

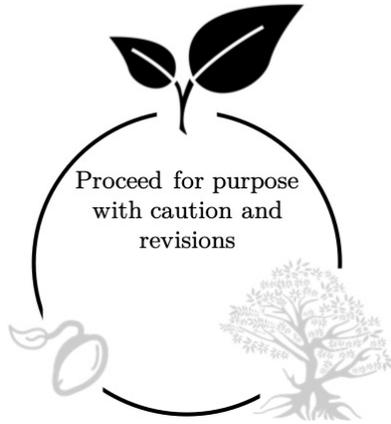
Draft September 2019

Abstract

Children living in war conditions frequently experience stressful conditions and multiple traumas, which can severely challenge their development, mental health, and academic functioning. It is a complex challenge to design education for areas with emergencies that meet local needs, are sensitive to local culture and context, build on international guidelines for best practice, and use research-based methods. In order to assess children's perceptions of the five domains of psychosocial support and care in the context of school-based psychosocial interventions, the Norwegian Refugee Council and the University of Tromsø worked to develop, pilot, refine, and test a measurement tool (PTS measure; version 1.0). Initial testing did not provide support for the hypothesized structure of the PTS measure. Exploratory factor analysis indicated an alternate five-factor structure with a reduced number of items, which was confirmed in a random half of the sample. Scores from the empirically derived subscales had mixed evidence of reliability and evidence of concurrent validity. Empirical analyses informed the development of the Student Learning in Emergency Checklist (SLEC-26; version 2.0), a tool that is developed for informing the work of planning, designing, and evaluating school-based psychosocial interventions for education in emergencies. The SLEC-26 helps measure promoters and barriers for learning before and after interventions to establish indicators for students' academic functioning and school well-being.

Correspondence concerning this report should be addressed to Dr. June T. Forsberg, e-mail: [june.t.forsberg@uit.no](mailto:june.t.forsberg@uit.no)

## Overview of SLEC-26: MENAT Measurement Library Criteria



SLEC-26 should have moderate to high evidence of validity/reliability for use as an evaluation measure, although it needs to meet less stringent criteria for use as a program monitoring tool with the goal of providing feedback to program implementers. While evidence did not support the original version of the measure (named PTS, version 1.0), there is promising evidence in support of the empirically derived structure of the SLEC-26 (version 2.0) measure. We are somewhat confident in the stability of the evidence, although the non-probability sampling may produce results that are unique to this dataset. We preliminarily recommend version 2.0 of this measure for its specified purpose(s) with attention to the additional recommended revisions. We also recommend additional testing of version 1.0 with attention to the recommended revisions.

Criteria	Indicators	Notes
<b>Purpose</b>	Program evaluation	Requires high internal consistency and interrater reliability; strong evidence of validity; sensitivity to change; ideally measurement invariance
	Program monitoring for improvement purposes	Requires less stringent criteria
<b>Empirical evidence overall</b>	# of types of evidence available	3
	% of evidence meets empirical criteria	60% (green only); 87% (yellow and green)
	Evidence fit for purpose	Yes for validity and reliability; not yet for measurement invariance/ sensitivity to change
<b>Confidence in evidence</b>	Sampling method	Multi-stage: <i>School level</i> : Non-probability; <i>Child-level</i> : All children present invited
	Sample size	Medium (~789)
	Missing data	Small amount of missing data
	Rigor of method	High
<b>Revisions</b>	Clear guidance on what to adjust/refine	Yes

## Overview of SLEC-26 Empirical Results: Version 2.0

Constructs/sub-constructs assessed	Internal consistency	Construct validity (internal structure)	Convergent validity	Recommendations for revision (if interested in using SLEC-26 constructs)
Sense of safety/adaptability	○	✓	✓	Consider narrowing and clarifying construct in next test of version 2.0
Emotion regulation	✗	✓	✓	Consider adding items to version 2.0 to improve reliability
School support	○	✓	○	
Family support	○	✓	✓	
Current and future well-being	✗	✓	✓	Consider narrowing and clarifying construct and adding items to version 2.0 to improve reliability

### Key

✓	Good/excellent evidence against empirical criteria	○	Fair/ inconclusive evidence against empirical criteria	✗	Little to no evidence against empirical criteria	NA	Not applicable
---	--	---	--	---	--	----	----------------

For additional information on the empirical criteria, please see <https://inee.org/measurement-library>

## Overview of PTS Empirical Results: Version 1.0

Constructs/sub-constructs assessed	Internal consistency	Construct validity (internal structure)	Recommendations for revision (if interested in using original PTS constructs)
Sense of safety	✗	✗	Consider dropping A3, adding to version 1.0 items similar to highest loading items
Emotion regulation	✗	✗	Consider dropping A10, adding to version 1.0 items similar to highest loading items
Self-efficacy	✗	✗	Consider dropping A14, adding to version 1.0 items similar to highest loading items
Social support	○	✗	Score separately as three separate constructs (school support, family support, peer support)
School functioning	✗	✓	Consider adding to version 1.0 additional items similar to highest loading items
Hope	✗	○	Consider adding to version 1.0 additional items similar to highest loading items

### Key

✓ Good/excellent evidence against empirical criteria	○ Fair/ inconclusive evidence against empirical criteria	✗ Little to no evidence against empirical criteria	NA Not applicable
--	--	--	-------------------

For additional information on the empirical criteria, please see <https://inee.org/measurement-library>



This technical working paper was developed by Drs. June T. Forsberg, Jon-Håkon Schultz, and Camilla Lodi as members of the 3EA | MENAT Measurement Consortium. Additional support was provided by Carly Tubbs Dolan. The report was reviewed by NYU Global Ties for Children.

Suggested citation: Forsberg, J., T., Schultz, J-H., Lodi, C., & Tubbs Dolan, C. (2019, August). *Development of the Student Learning in Emergency Checklist (26) (SLEC-26): A measurement tool of promoters and barriers for learning among conflict-affected students in Palestine*. Technical working paper. Tromsø, Norway: University of Tromsø.

## Introduction

In the context of crisis, disasters, or war, school-aged children are a vulnerable group that have access to fewer coping resources than adults. The school is viewed as a key arena for preventive interventions to reduce negative reactions and stress and improve school functioning. Teachers can be given an important role in large-scale school-based interventions following war or during on-going conflict. Teachers, with the support of parents, have an important role in supporting children to stabilize and recover, so they can learn and thrive.

A systematic review of 83 studies has shown that children with exposure to recurrent and/or severe traumatic events were at significant risk for impairments in cognitive functioning, academic difficulties, and social-emotional-behavioral problems (Perfect, Turley, Carlson, Yohanna, & Saint Gilles, 2016). In addition, the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) draws on a strong foundation of studies on traumatic stress when describing the categories of symptoms potentially related to traumatic stress: behavioral problems, impaired cognitive functioning, somatic distress, and psychological reactions.

A comprehensive review of intervention research on the treatment of those exposed to disasters and mass violence have identified five widely accepted and empirically supported principles that are used to inform intervention and prevention efforts, both in the immediate aftermath of a critical event and up to three months thereafter (Hobfoll et al., 2007). These five principles are: (1) to promote a sense of security, (2) to calm, (3) to foster a sense of self- and collective efficacy, (4) to promote connectedness, and (5) to instill hope. These principles are included in guidelines such as Psychological First Aid (PFA; Brymer et al., 2013), “The European Network for Traumatic Stress Guidelines” (Bisson et al., 2010), and Skills for Psychological Recovery (SPR; Berkowitz et al., 2010). The guidelines present best practices for mental health and psychosocial support (MH/PSS) after a critical event. The five essential principles are used in school-based delivery of post-disaster psychosocial care (e.g., Psychological First Aid) and also in war and on-going conflict (e.g., The Better Learning Program, BLP: NRC, 2018).

In order to assess children’s perceptions of the five domains of psychosocial support and care in the context of school-based psychosocial interventions, the Norwegian Refugee Council and the University of Tromsø worked to develop, pilot, refine, and test a measurement tool. The tool was intended to be used to (1) establish a baseline; (2) provide information about the target group for course correction purposes (*program monitoring for feedback within schools and to program implementers*); and (3) measure change in the five domains after the intervention is carried out (*program evaluation*). The tool was developed in stages, including through an initial pilot study that examined the relationship between Gaza students’ self-reported school functioning (academic functioning and school well-

being) and the five domains of psychosocial care and support (see Forsberg & Schultz, in prep). The measure was then further revised and a psychometric testing study of the measure (hereafter PTS measure; also version 1.0 in Library) was conducted. For an in-depth description, see Forsberg, Dolan, and Schultz (in prep). Analysis of the psychometric results of the PTS measure resulted in a new version of the measure – SLEC-26 (version 2.0) -- for Palestinian students in the age range of 12–16 years old. Below we provide an overview of the psychometric study and the description of the results that led to the SLEC-26.

## Method

### Participants/sample

Eight hundred twenty-five Palestinian school students were recruited and participated in the psychometric testing study. Thirty-six students were excluded from the sample because they were not within the age range of 12–16 years old.<sup>1</sup> This resulted in a sample of 789 students (358 males and 431 females).

The sample was geographically recruited from the Gaza strip (N = 397), Jerusalem (N = 150), Hebron H2 (N= 114), or other areas of the West Bank (N= 126). Two students did not report where they belonged geographically. The students were recruited from a total of 17 schools, grade 5-10 (9 boys schools, 5 girls schools and 3 mixed schools) that had implemented the Better Learning Program (BLP) intervention and were frequently supervised by the Norwegian Refugee Council (NRC). The schools were selected due to demographic characteristics to be representative of the universe of schools implementing BLP. Geographic areas within Gaza (North, East and West), Jerusalem, Hebron H2, and other areas of the West Bank were equally represented. In total there are 55 schools in Gaza (44 schools are schools of the Ministry of Education in Gaza (MoE) and 11 schools are schools of the United Nations Relief and Work Agency (UNWRA)) and 78 schools in West Bank (all MoE schools) that have implemented BLP. The schools were all basic schools from grades 1-10, and they had an average of 595 students per school.

All the students that were present on the days that NRC visited the schools for the purpose of the study were invited to participate in the study. All classrooms in grades 7-10 were visited. Since all the schools were already enrolled in BLP, the parents had provided written consent that the children could participate in research surveys about the program. Participation in this specific psychometric testing study was voluntary and the students gave their informed oral consent for participation. Before they consented, they were informed that the purpose of the study was to learn more about the best way for students to function in school, and that the data would be used in a research project. Furthermore, they were informed that all information would be private and kept anonymous. The

---

<sup>1</sup> One student was 11 years old, 14 students were 17 years old, 4 students were 18 years old, and one student was 19 years old. In addition, 17 students were excluded because they did not report their age in the questionnaire.

Palestinian Health Research Council and the Norwegian Centre for Research Data approved the study. MoE approved the study protocol and procedure.

### **Measure**

The tool that was used in the PTS to develop SLEC-26 was a self-report questionnaire consisting of 6 sections (A-F) and a total of 67 items. The sections were designed to assess seven different domains, five of which were based on the principles presented by Hobfoll et al. (2007). In addition, we also included school functioning and traumatic stress as domains. Section A included 26 items about the feeling of safety, self-regulation, self-efficacy, and social support. Section B assessed academic functioning and school well-being, in total four items. Section C was an exposure scale that assessed traumatic events. Section D consisted of the Revised Child Impact of Event Scale (CRIES-13), a standardized tool to assess traumatic stress after stressful life events in children. Section E included five items about the feeling of hope and section F mapped demographic characteristics.

All items in sections A, B, and E were statements that the students scaled on a 1-5 Likert scale (always, most of the time, sometimes, rarely, or never). Section C (the exposure scale) listed 13 traumatic events that are typical for the Palestinian contexts, and the students marked the events that they had been exposed to in the last three months. Next, the students marked which traumatic event scared them the most and scaled from 1-5 how afraid the event made them. In section D (CRIES-13), the items were formulated as questions with four different response options (not at all, rarely, sometimes, and often). Since the CRIES-13 is a standardized tool, no changes were made to the content. Section E, about the feeling of hope, was presented at the end of the questionnaire with the intention that the students would finish the survey in a positive and hopeful state of mind.

**Development process.** The tool that was used in the PTS was developed over a period of five months prior to the data collection in February 2019. The process started with a two-day workshop in September 2018 in Tromsø, in which both The University of Tromsø (UiT) and NRC participated. The purpose of the workshop was to conceptualize the domains for the tool and plan an evaluation study (the pilot study) of the BLP in Gaza. The purpose of the pilot study was to test the function of the tool in the correct context and similar sample as the PTS. Since the pilot study was conducted in Gaza, the first focus was to develop a tool for the Gaza context. After the workshop in Tromsø, UiT designed the tool and prepared items, before NRC was consulted. The tool was then revised four times over the next two months, in October and November 2018.

The items in sections A, B, and E were drawn from and/or inspired by earlier evaluation surveys that NRC had used to evaluate the effect of BLP. The items were developed by Prof. Schultz and targeted to evaluate the principles of Hobfoll et al. (2007) on which BLP is based.

In the revising process, four main areas were targeted:

- (1) **Number of items in the tool:** The tool needed enough items to cover all the domains, but also be short and manageable for the students. To secure validation it was important to arrange the tool so the students could keep concentration throughout the survey.
- (2) **Wording:** The language in each item was specifically adapted to target children in Gaza in the age range of 12–16 years old.
- (3) **The political situation and ongoing conflict in Gaza:** NRC gave input on how to formulate the items in the context to facilitate approval by the MoE in Gaza.
- (4) **The exposure scale:** The purpose of the exposure scale was to map traumatic events that the students had been exposed to. NRC staff in Gaza suggested traumatic events that were relevant in the Gaza context.

This resulted in a 30-item questionnaire that was translated to Arabic and NRC conducted a small research pilot (pre-pilot test) in October 2018, which included 50 students in five Gaza schools. Based on the feedback from the students and the teachers, a small revision in the Arabic language was conducted. The pilot study (N = 250) was then conducted in 25 schools in November and December 2018. During this period, UiT and NRC visited the schools and gathered experiences from the students and the teachers about the tool. They were interviewed about the design, wording, and number of items. They also gave input on what they liked about the tool and what they found challenging. The interviews were non-structured interviews between UiT, the schools' principals, and teachers/counselors that provided the BLP intervention in the classroom. NRC staff were also present and translated between English and Arabic. The interviews were recorded with a cellphone.

The pilot study and the interviews gave valuable information about the function of the tool, and a fifth revision was conducted in January 2019. In the fifth revision, two main areas were targeted:

- (1) **Number of items in the tool:** In the previous revisions, a main focus was that the tool needed to be as short and manageable as possible. In the fifth revision this focus changed. For the purpose of conducting a factor analysis, more items were included for each domain in section A, B, and E. A total of five items per domain were included in each section. The additional items were also drawn from earlier BLP evaluation surveys, developed by Prof. Schultz.
- (2) **The exposure scale:** Since the ongoing situations in Gaza and the West Bank are very different, the exposure scale was extended to include the West Bank context as well. To develop the extended exposure scale, we arranged focus group meetings with NRC staff in Gaza and in the West Bank and we also invited teachers and psychological counselors that worked in the schools where BLP was enrolled. The teachers and the counselors provided valuable and insightful suggestions for traumatic events that applied for both Gaza and the West Bank.

After the fifth revision, we had a 67-item tool for the PTS. A native Arabic-speaker from the NRC staff translated the questionnaire to Arabic and then we hired a free native English-speaking agent to translate the tool back to English again. The back-translation revealed that three items needed wording changes in Arabic so that they would match the English version and sharpen the original intension of the items. In February 2019, the NRC administered the tool in the selected schools in Palestine for the purpose of developing SLEC-26. The development process is illustrated in Figure 1 and the PTS tool is presented on the Library website.

## **Analytic Plan**

In this section, we present an overview of the methods and analyses that were used to assess the function of the tool and the analyses that resulted in SLEC-26.

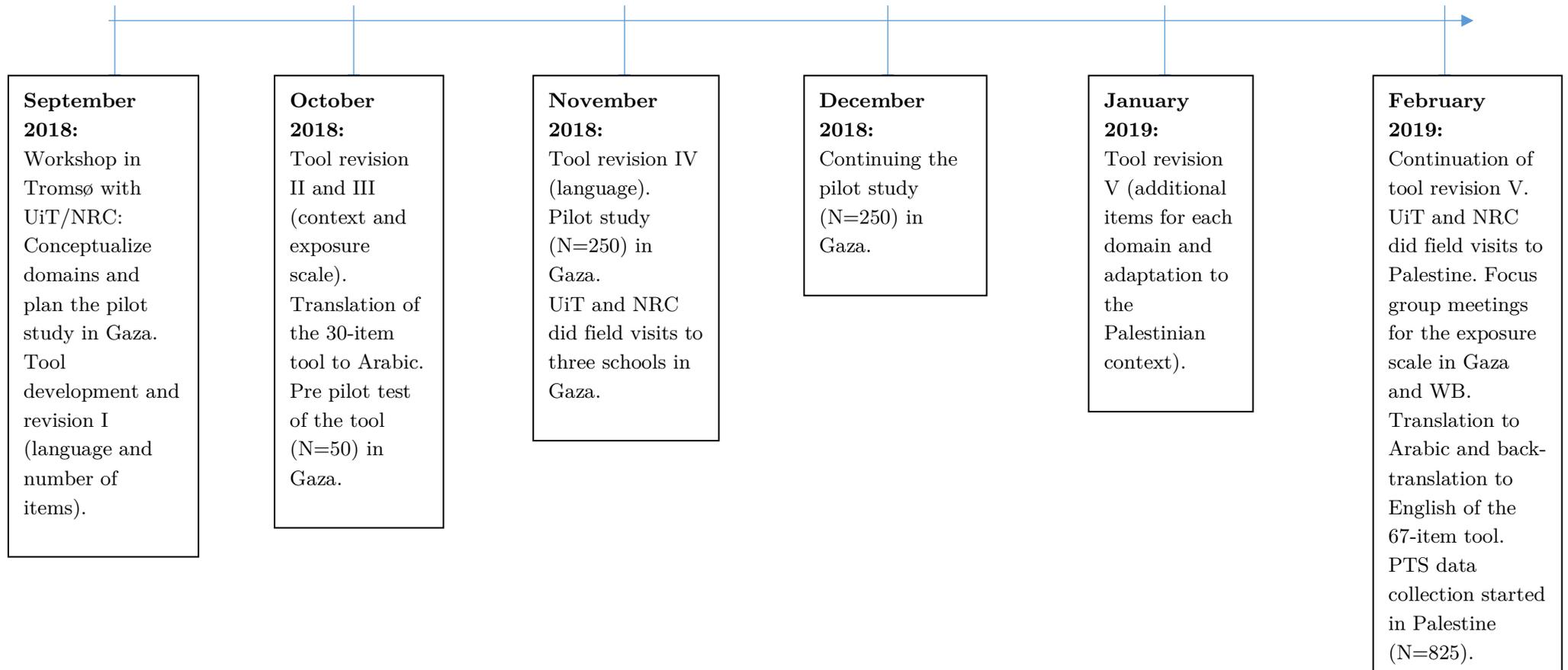
### **Content validity**

Content validity assessment was a continuous process throughout the development of the PTS tool (later SLEC-26), and the purpose was to ensure correct functioning of the items in the tool. The content validity assessment was performed for each section separately. Sections A and E included items that measured the five principles of Hobfoll et al. (2007). We focused on the different theoretical constructs (domains) on which the principles are based and designed the items to capture the constructs. Items to measure self-efficacy were retrieved from the General Self-Efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995). The other items were designed by Forsberg and Schultz and drawn from and/or inspired by earlier evaluation surveys that NRC had used to evaluate the effect of BLP.

Section B measured school functioning, which included self-reported academic functioning and well-being. The items in section B were designed by Forsberg and Schultz and were based on the pedagogical principles in BLP. Sections C and D measured exposure to traumatic events and stress level. The exposure scale was designed by Forsberg and Schultz and the items in section D were retrieved from the Children's Impact of Event Scale (13), CRIES-13.

All items were designed to capture the different domains. Next, the items were piloted on the target population in the Gaza context. After the pilot we performed qualitative interviews with teachers and students to get feedback on the function of the items. The items in the exposure scale were assessed in focus group meetings together with NRC staff from Gaza and the West Bank. However, the exposure scale and CRIES-13 are not included in SLEC-26 but can be used as appendixes to explore traumatic stress.

Figure 1. The development process of the PTS tool.



## Construct validity

**Internal structure.** In order to assess the internal structure of sections A, B, and E of the measure, we conducted a series of exploratory and confirmatory factor analyses using MPlus version 7.2 (Muthén & Muthén, 2014). In order to account for structural characteristics of the data, two important specifications are made. First, students' responses were assessed on an ordinal 5-point scale in which the assumption of equal increments between scale points cannot be made. Because modeling the responses as normally and continuously distributed can lead to 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, 2009). Second, as described above, students in this sample were nested in schools, thereby violating the assumption of independence of standard errors required in the application of factor analytic techniques. We thus used the TYPE=COMPLEX command in MPlus to estimate robust standard errors (Muthén & Muthén, 2012). Factor analyses proceeded in three steps:

**Step 1: Randomly split sample.** We first randomly divided the full sample in half stratified on gender and grade in order to create exploratory and confirmatory samples. 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 hypothesized and proposed factor structures, thereby building confidence in the stability of empirically derived exploratory factor analytic estimates (Osborn & Fitzpatrick, 2012).

**Step 2: Conduct exploratory analyses by hypothesized domain.** Using the exploratory sample, we first tested whether there was evidence that correlations among items within each hypothesized domain of sections A, B, and E were consistent with a single latent construct. We did so by first fitting one-factor confirmatory models for (1) perceived sense of safety; (2) emotion regulation; (3) self-efficacy; (4) social support – school, family, and peer; (5) school functioning/well-being; and (6) future hope. To assess the goodness of fit of the models, the following two criteria were used (Hu & Bentler, 1999): (a) a root mean squared error of approximation (RMSEA) value below .08 provides an acceptable fit to the data, while an RMSEA of less than .05 provides a good fit to the data; and (b) a comparative fit index (CFI) value above .9 provides an acceptable fit to the data while a CFI value above .95 provides a good fit to the data (Kline, 2011).

**Step 3: Conduct exploratory factor analyses to identify empirically derive subscales and confirm the solutions out-of-sample.** As discussed in the results section below, while some of the confirmatory models provide a good fit to the data, many of the resulting subscales do not have evidence of reliability. We thus conducted an exploratory factor analysis using items from sections A, B, and E to examine whether it was possible to empirically derive subscales that aligned with theory but resulted in scores

with better reliability. Analyses were conducted using the exploratory sample with an oblimin rotation. Based on a combination of the overall goodness-of-fit statistics, item specificity and strength, targeted misspecification indices, and face validity of the models, we selected a final factor structure to test using the confirmatory sample.

**Concurrent validity.** To assess concurrent validity, we calculated sum scores for each of the empirically derived subscales. Such a method makes the assumption that all items have equal weight in the composite scores; however, unit-weighted scores are the most common scoring method used by practitioners in humanitarian contexts. As such, we examine zero-order Pearson correlation coefficients between the empirically derived sum scores, children’s demographic information, and other theoretically related constructs., including cumulative exposure to war events, and traumatic stress responses (sections C and D of the PTS measure)<sup>2</sup>.

### **Reliability**

Internal consistency for the hypothesized and empirically derived subscales was tested with Cronbach’s alpha.

## **Results**

Below we share results of descriptive item-level analyses as well as provide a brief overview of the results of efforts to (1) confirm the hypothesized factor structure of the measure; and (2) empirically explore the factor structure to derive a set of subscales with evidence of validity and reliability. Detailed results from the analysis are presented in Appendix 1, which is available upon request from the developers. For detailed results from the analyses, see Forsberg et al., in prep.

### **Descriptive statistics**

Item-level means, standard deviations, minimums and maximums for items from sections A, B, and E of the PTS measure are presented in Table 1. On average, Palestinian students reported feeling safe on the way to and at home and school most of the time. However, students reported feeling calm at school only sometimes. In general, students reported sometimes being able to control feelings of sadness, fear, and anger. They also reported being able to manage problems, to stick to their aims, and to adapt to unpredictable circumstances sometimes to most of the time. In terms of social support, students said that school staff only rarely to sometimes asked how they were doing or provided support; students reported perceiving higher levels of support from family and friends. Students also responded on average that they were sometimes to most of the time able to concentrate and do their best at school. Finally, Palestinian students reported feeling hopeful for the future and future events most of the time.

---

<sup>2</sup> Factor analyses of the measures of exposure to traumatic events and stress response are not presented in the current report but are available on request.

Table 1. *Item-level descriptive statistics*

Items	<i>N</i>	Mean	SD	Min	Max
Perceived Safety ( $\alpha = 0.55$ )					
I feel safe at school	783	4.10	1.06	1	5
I feel that the teachers and school staff respect me	782	3.81	1.17	1	5
I feel calm at school	784	2.91	1.25	1	5
I am afraid when I am at school (Reverse)	784	4.40	1.03	1	5
I am scared to travel to school or home from school because the trip can be dangerous (Reverse)	784	4.36	1.11	1	5
I feel safe at home	786	4.75	0.76	1	5
Emotion Regulation ( $\alpha = 0.67$ )					
When I feel sad, I know what to do to feel better	785	3.59	1.24	1	5
I can control my emotions	785	3.02	1.32	1	5
When I am scared, I can calm myself down	780	3.50	1.26	1	5
When I get angry, I hit other people or things (Reverse)	777	3.70	1.36	1	5
When I feel angry, I can calm myself down	782	3.34	1.25	1	5
Self-efficacy ( $\alpha = 0.67$ )					
I can manage to solve difficult problems if I try hard enough	782	3.82	1.11	1	5
It is easy for me to stick to my aims	784	4.18	1.08	1	5
It is easy for me to accomplish my goals	786	3.78	1.12	1	5
If I am in trouble, I can think of a solution	786	4.02	1.11	1	5
I can handle whatever comes my way	786	3.68	1.12	1	5
I am confident that I can adapt to events I had not predicted	781	3.31	1.30	1	5
Social Support ( $\alpha = 0.68$ )					
I can talk to someone on the school staff (a teacher, the principal, a counselor) about my worries	783	2.63	1.53	1	5
Someone on the school staff (a teacher, the principal, a counselor) asks me how I am doing	783	2.62	1.44	1	5
Someone on the school staff (a teacher, the principal, a counselor) supports me when I feel scared	779	2.68	1.49	1	5
I can talk to my parents about my worries	779	3.80	1.42	1	5
When I feel scared, I can tell my parents	778	3.75	1.43	1	5
My parents ask me how I'm doing	782	4.56	0.91	1	5
I have friends I like to play with at school	780	4.53	1.00	1	5
I can talk to my friends about my worries	782	3.17	1.44	1	5
My friends support me when I feel scared	785	3.67	1.35	1	5
School Functioning ( $\alpha = 0.63$ )					
I can easily concentrate when doing school work	783	3.76	1.15	1	5
I am able to do my best in school	786	3.98	1.11	1	5

I am satisfied with my life	786	4.06	1.26	1	5
I like being at school	786	3.30	1.40	1	5
Sense of Hope ( $\alpha = 0.62$ )					
Things will turn out well in the future	772	4.09	1.16	1	5
I will graduate school	782	4.35	1.22	1	5
I will get a job when I grow up	784	4.28	1.11	1	5
I will have a family when I grow up	782	4.26	1.25	1	5
I will live a meaningful life when I grow up	783	4.47	0.99	1	5
Demographics					
Female	789	55%		0	1
Age	789	14.29	1.29	12	16
Grade level	787	8.66	1.20	7	10
Gaza	787	54%		0	1
Jerusalem	787	19%		0	1
Hebron	787	14%		0	1
Other West Bank	787	16%		0	1

### Construct validity

**Confirmatory factor analysis by hypothesized domain.** We then conducted a series of confirmatory factor analyses to understand whether the intercorrelations among items within each hypothesized domain were consistent with a single latent factor. Model fit statistics for the one-factor models are presented in Table 2, below.

Table 2. *Summary of goodness of fit statistics for hypothesized one-factor models*

	$\chi^2$ (df)	RMSEA (90% CI)	CFI
Sense of safety	25.712 (9)	0.069 (.038-.101)	0.862
Emotion regulation	26.513 (5)	0.104 (.067-.145)	0.955
Self-efficacy	52.130 (9)	0.110 (.082-.140)	0.892
Social support	321.945 (27)	0.166 (.150-.182)	0.605
School functioning	3.061 (2)	0.037 (0.00-.113)	0.994
Sense of hope	18.134 (5)	0.082 (.043-.123)	0.948

While correlations among items in the hypothesized school functioning and sense of hope scales were consistent with a single latent construct, a one-factor model did not provide a good fit to the data from the hypothesized sense of safety, emotion regulation, self-efficacy, and social support items. In addition, as shown in Table 1, the internal consistencies of all scores calculated according to the hypothesized scale structure was below 0.70, the commonly used accepted cutoff value for alpha (Yang & Green, 2011).

**Exploratory factor analyses.** Using a random half sample of the data, two types of exploratory factor analyses were conducted. First, in order to understand whether a

reduced number of domains could potentially represent the items in the PTS measure, we examined the scree plot; looked for eigenvalues  $> 2$  cut-off criteria; and examined the model fit statistics of 1- through 7-factor exploratory factor analytic solutions. Based on these results (available upon request from the developer) as well as examination of the factor structure at the item level, a five-factor structure emerged:

Table 3. *Empirically derived factors, construct names, and items*

Factor	Constructs	Items
	Sense of safety and adaptability	A1, A2, A6, A12, A14, A15, A16, A17, B1
	Emotion regulation	A8, A9, A10, A11
	School support	A18-20
	Family support	A21-23
	Present and future well-being	B28-30, E59, E60, E61, E63

Cross-loading and low-loading items suggested by the exploratory analysis were removed before confirming the final proposed model.

We also conducted a set of exploratory analyses within each originally hypothesized domain to understand whether model fit could be improved by removal of certain items with low factor loadings or high model residual correlations. While these modifications likely would not improve the internal consistency of sub-scale scores, they can guide future revisions of the measure if the user is specifically interested in measuring the six core constructs hypothesized to be assessed in the PTS measure.

Table 4. *Summary of goodness of fit statistics for revised one-factor models*

	$\chi^2$ (df)	RMSEA (90% CI)	CFI	Modification
Sense of safety	6.837 (5)	0.030 (.00-.081)	0.992	Remove A3
Emotion regulation	5.48 (2)	0.066 (.00-.135)	0.995	Remove A10
Self-efficacy	6.61 (5)	0.028 (.00-.080)	0.993	Remove A14
Social support	21.003 (12)	0.043 (.00-.074)	0.988	Model as 3 factors
School support				Include items A18-20
Family support				Include items A21-23
Peer support				Include items A24-26
School functioning	3.061 (2)	0.037 (0.00-.113)	0.994	
Sense of hope	.758 (2)	0.000 (.00-.075)	1.00	Remove E62

**Confirmatory factor analysis of empirically derived subscales.** Based on the results of the exploratory analyses, two sets of confirmatory factor analyses were conducted with the other random half of the sample.

First, the CFA model with the five main factors identified through the EFA (see Table 3, above) provided a good fit to the data ( $\chi^2$  (288) = 359.60,  $p = 0.003$ , RMSEA (90% CI) = 0.025 (0.016-0.033), CFI = 0.946). Table 1 in Appendix 1 (available upon request from

developers) presents the standardized factor loadings of items, as well as correlations among factors. Items showed moderate to high ( $>0.40$ ) loadings on their factor, with the exception of A10. This item was a negatively worded item that was reverse-coded; future revisions may want to change the wording of this item.

Correlations between factors were all in the expected, positive direction. Correlations were low to moderate, with the exception of a high correlation between the safety and adaptability factor and the current and future well-being factor. This suggests that these factors provide a high degree of overlapping information, and future revisions of the measure should consider reviewing and adjusting the items in those factors to further clarify and distill the constructs.

Second, a CFA model with the eight factors identified by the EFAs of the individual hypothesized domains (see Table 4) was also fit to the data. This model also provided a good fit to the data ( $\chi^2(405) = 475.80, p = 0.008, RMSEA(90\% CI) = 0.021(0.011-0.029), CFI = 0.949$ ). Items showed moderate to high ( $>0.40$ ) loadings on their factor, with the exception of A4 and A5. These items were negatively worded item that were reverse-coded; future revisions may want to change the wording of this item. Correlations between factors were all in the expected direction, and correlations were low to moderate, with the exception of the correlation between safety and school functioning ( $r = .81$ ).

### **Reliability**

Cronbach's alpha coefficients of the subscale sum scores were calculated from the empirically derived five- and eight-factor models, respectively. The reliabilities of three (safety and adaptability, school support, family support) of the five subscale sum scores calculated according to the empirically derived five-factor model are above 0.70, the commonly accepted cutoff value for alpha (Yang & Green, 2011; see Table 3, Appendix 1). While this is an improvement in reliability over the scores from hypothesized six-factor model, emotion regulation and current and future well-being sum scores remain below the commonly accepted criteria. In addition, as expected, the reliabilities of most of the subscale sum scores calculated according to the empirically derived eight-factor model are below 0.70, with the exception of scores from school and family support scales (see Table 4, Appendix 1).

Given that confirmatory factor analytic fit indices of the five- and eight-factor models were acceptable, there is conflicting evidence about the extent to which the empirically derived scoring strategy is supported by the data. Such a situation can arise when (a) scales are too short, given that calculation of coefficient alpha is dependent on the number of items in the scale; and (b) the underlying construct is defined and operationalized broadly, as a consequence of which item responses are less closely linked. We thus recommend that future iterations of either the SLEC-26 (version 2; the five-factor model) or of the original PTS tool (version 1) consider clarifying and narrowing certain constructs (i.e., safety, self-efficacy) and adding additional items (e.g., school functioning, hope) to ensure reliability.

## **Concurrent validity**

**Correlations with age, gender, mental health, and exposure to war events.** We first examined bivariate correlations between empirically derived SLEC-26 constructs and children's gender, age, and grade using sum scores of each construct. (See Table 1 in Appendix 1, available upon request from developer). There were two significant and moderate correlations between gender and SLEC-26 constructs: Girls perceived less support from school staff than boys, but reported greater perceived current and future well-being. Correlations between SLEC-26 constructs and age, and between SLEC-26 constructs and grade followed the same pattern and are reported here together. In general, the magnitude of the correlations was small but significant. Older children/children in higher grades reported less frequently feeling safe and able to adapt to difficulties, and reported lower perceived well-being. Older children/children in higher grades also perceived less support from school staff and family.

Correlations of SLEC-26 constructs with PTSD symptoms and exposure to war events were small, significant, and generally in the expected direction. Youth who reported higher levels of PTSD symptomology and youth who reported greater cumulative exposure to war events reported less frequently feeling safe and able to adapt to difficulties; greater difficulty regulating their emotions; and lower perceived current and future well-being. They also perceived less support from their family. Interestingly, however, there was not a significant association between PTSD symptomology and perceived support from school staff, and youth who reported greater cumulative exposure to war events perceived significantly greater support from school staff.

## **Limitation**

The two biggest challenges and limitations in this process have been (1) the distance and context differences between Tromsø, Norway and Palestine and (2) doing research in an ongoing conflict.

(1) The distance and context differences between Tromsø and Palestine were challenging in several steps of the project. It is more challenging communicating via e-mail and skype than face to face. It reduces the communication frequency and it increases the threshold of having misapprehensions and making mistakes on both sides. Nevertheless, establishing clear areas of responsibility was very helpful and there have been no serious mistakes in the methodological process. The context differences were challenging during the tool development and adaptation, and also during the data collection. It made the processes very time-consuming, and both UiT and NRC had to make several compromises to adapt the scientific methodical work to the complex practice reality in Palestine.

(2) Doing research in areas with ongoing conflict is challenging because there is no way to maintain scientific and quality control in every step the methodological approach. The participants are also affected by the ongoing conflict, which can influence how they respond in the survey. In November 2018, during the data collection for the pilot study,

there were a series of bombing in Gaza. We did not find that the bombing influenced the effect of BLP (see Forsberg & Schultz, in prep), but it could have influenced how they responded in the survey and how they considered the items. During the data collection for the PTS, it was a turbulent period all over Palestine, which can be considered a limitation of the study.

## **Discussion**

The analysis revealed a different factor structure than what we expected, which resulted in adjustments in the categories and a new version of the PTS measure: SLEC-26. Since the overall factor model was good and most of the reliabilities were acceptable or near acceptable, SLEC-26 satisfies the monitoring purposes for the tool. The internal consistencies are generally too low for the evaluation and research purposes of the tool. To secure these purposes in the future as well, we need to revise SLEC-26 before it can be sufficiently validated and standardized.

A revision of SLEC-26 would imply a new data collection with the PTS-tool in the same population as in this study. The PTS-tool would be revised and items with low factor loadings would be exchanged with new items in each of the domains. Domains could be selected from either the five-factor model represented in SLEC-26 or from the empirically derived eight-factor model reported here, depending on what constructs the user is interested in assessing. The content of the new items should be similar to already existing items with high factor loadings. This should improve the internal consistency in all categories/subscale in SLEC-26 or the eight-factor model.

### *Standardizing SLEC-26 to another population and context*

The measure is designed for use with the specific context of students of both genders, between the ages of 12–16, living in Palestinian occupied territory: the Gaza strip and the West Bank. The process of adapting the measure for other populations and contexts would need comprehensive work. Such work would entail considering the age, gender, and background (e.g., religion, refugee status, country/area, and cultural context) of the students.

If a comprehensive standardization process is feasible, we recommend following a common systematic adaption approach to ensure that SLEC-26 is contextually relevant outside of Palestine.

#### **1) Review by national staff:**

Field and program staff that are experienced in the relevant context should review the tool and assess the content. They need to assess if the wording in the items is appropriate for the age group that is targeted. In addition, they need to consider cultural and contextual requirements.

#### **2) Translation:**

The tool needs to be translated from either English or Arabic to the native language of the students that are targeted.

**3) Review of translation:**

After the tool is translated, it must be reviewed by national staff as in point 1).

**4) Back-translation:**

Back-translation to English or Arabic is needed to validate the translation and to ensure that the intention and meaning of the items are correct.

**5) Pre-testing assessment:**

A pilot test should be conducted using the tool on a small and similar sample as the students that are targeted, and feedback from the students and the staff that administer the tool should be reviewed.

**6) Finalization:**

Make adjustments according to the feedback and consider if a new translation process is necessary.

## **Appendix**

- 1) Detailed results from the analysis. Available upon request from developer

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Berkowitz, S., Bryant, R., Brymer, M., Hamblen, J., Jacobs, A., Layne, C., & Watson, P. (2010). *Skills for psychological recovery: Field operations guide*. Los Angeles: National Center for PTSD and National Child Traumatic Stress Network.
- Bisson, J. I., Behrooz, T., Witteveen, A., Ajdukovic, D., Jehel, L., Johansen, V.J., . . . Olf, M. (2010). TENTS guidelines: Development of post-disaster psychosocial care guidelines through a Delphi process. *British Journal of Psychiatry*, *196*(1), 69–74.
- Brymer, M., Jacobs, A., Layne, C., Pynoos, R., Ruzek, J., Steinberg, A., . . . Watson, P. (2013). *Psychological first aid: Field operations guide* (2nd ed.). Los Angeles: National Child Traumatic Stress Network and National Center for PTSD.
- Forsberg, J. T., Dolan, C.T., & Schultz, J-H. (in prep). The student learning in emergency checklist (SLEC-26): A measurement tool of promoters and barriers for learning among conflict affected students in Palestine.
- Forsberg, J. T., & Schultz, J-H. (in prep). Helping students living in ongoing conflict: The effect of a school-based intervention program targeting academic underachievement.
- Hobfoll, S. E., Watson, P., Bell, C. C., Bryant, R. A., Brymer, M. J., Friedman, M. J., . . . Ursano, R. J. (2007). Five essential elements of immediate and mid-term mass trauma intervention: Empirical evidence. *Psychiatry: Interpersonal and Biological Processes*, *70*(4), 283–315.
- Perfect, M. M., Turley, M. R., Carlson, J. S., Yohanna, J., & Saint Gilles, M. P. (2016). School-related outcomes of traumatic event exposure and traumatic stress symptoms in students: A systematic review of research from 1990 to 2015. *School Mental Health*, *8*(1), 7–43. <http://doi.org/10.1007/s12310-016-9175-2>
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor, UK: NFER-NELSON.