Testing Measures of Refugee Camp Environment, Caregiver Mental Health, and Child Social-Emotional Development Among the Rohingya in Cox’s Bazar

Yeshim Iqbal, Sneha Bolisetty, Ahmed Alif, Priyamvada Tiwari, Dennis Hilgendorf, and Hirokazu Yoshikawa

SEPTEMBER 2022
Introduction

As a part of the Play to Learn initiative\(^1\), New York University’s Global TIES for Children, along with Bangladesh Rehabilitation Assistance Committee (BRAC) and Innovations for Poverty Action (IPA), collected pilot data in the Rohingya camps in Cox’s Bazar district, Bangladesh, with the aim of developing culturally sensitive measures for assessing child development and caregiver well-being. The Rohingya are an understudied population with a unique linguistic and cultural history, and as such few survey measures exist that have been validated in areas that are important to early childhood development. Many measures commonly used to assess children’s development and immediate environment in low and middle income country contexts\(^2\)\(^3\)\(^4\)\(^;\)\(^5\)\(^6\)\(^7\)\(^;\)\(^8\)\(^9\), and in humanitarian contexts\(^5\)\(^6\)\(^7\)\(^;\)\(^10\)\(^11\)\(^12\); have not been previously used in the Rohingya context. Given the lack of culturally sensitive measures for this group that are relevant to early childhood development, the goal of the study outlined in this brief was to assess survey measures relevant to children’s development specific to the Rohingya refugee camps. Specifically, we measured experiences of the refugee camps, caregivers’ well-being, and child social-emotional development among families with pre primary-aged (3-4 year old) children.

In this brief, we provide a snapshot of the process we went through to test scales\(^10\) in this context. We outline the steps involved in selecting, adapting, and testing these scales to examine their suitability and prepare them for large-scale use. We then present analyses of these scales, including factor structure, reliability, validity, and scale inter-correlations. We conclude with a summary of what we learn from our results and future directions.

Background

Over 900,000 Rohingya people live in refugee camps in the Cox’s Bazar region of Bangladesh\(^13\), and of them almost 500,000 are children\(^14\). For these children, statelessness and lack of legal status create difficulties in accessing basic services, including education\(^15\). Although a substantial number of children are accessing early childhood education through programs such as BRAC’s Humanitarian Play Labs (HPLs), early childhood instruments, which measure the cognitive, language, and social-emotional development of children and their immediate environments, have not been previously adapted or tested for cultural relevance in the Rohingya context\(^16\). Development of usable instruments is critical to assess the impact of different interventions for Rohingya children, with substantial implications for improving practice and policy.
Constructs of interest

We focus in this brief on experiences of the refugee camps, caregiver mental health, and child social-emotional development among Rohingya families. Here, we briefly indicate why these constructs are important and especially relevant to understanding early childhood development in the humanitarian context.

Caregivers’ perceptions of the refugee camp environment

Studies indicate that both pre- and post-migration experiences can powerfully shape refugee and displaced families' and children’s development. Yet experiences of the refugee camp context – access to basic services; housing adequacy in the face of unpredictable climate and other conditions; food insecurity; and the human context of available social support – have rarely been linked to young children’s development. Caregivers’ experiences of the everyday stresses and supports in these generally turbulent environments may affect their well-being and in turn their children’s development.

Caregiver mental health

Here, we assess caregiver (in this case, maternal) mental health in the form of depressive symptoms and stress. These may powerfully affect caregivers’ general functioning and children’s social-emotional development. Although maternal depression has been linked to infant physical health and temperament in rural Bangladesh, it has not been examined with reference to pre primary-age children nor with the Rohingya. Similarly, stresses related to the parenting role are an aspect of well-being that is specific to socialization of young children. Parenting stress in Bangladesh has been examined only with regard to parenting children with disabilities or illness, and no studies exist that have assessed parenting stress among Rohingya caregivers.

Children’s social-emotional development

This is a key dimension of early childhood development, but may manifest differently in different cultural contexts. Social behaviors include both negative aspects like acting-out or withdrawn behaviors, as well as positive aspects such as prosocial, cooperative, or helping behaviors. Again, relatively few studies have been conducted in the current region and none with the Rohingya. In a recent systematic review of social-emotional development in Asia, for example, only a few studies were found that focused on South or Southeast Asian contexts.
Selected measures for constructs of interest

Measures were initially selected through reviews of research literature and revision from the NYU research team. Generally we were interested in scales that have been previously used and validated (especially those that have been used in LMIC or humanitarian contexts) and that have commonly been used for research in early childhood development. Once a particular scale was selected, its individual items were examined closely; in some cases, the scale was kept with all items even if we were unsure about certain items’ utility in the context, in the interest of keeping an original scale whole. In other cases, particular items were dropped at the outset because they were deemed definitively inappropriate for the context.

Table 1 below lists the selected measure for each construct, with example items, response types, and a brief description of the measure’s respective previous use.

Table 1. Selected measures for assessing caregiver perception of environment, caregiver mental health, and child social-emotional development.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Description</th>
<th>Item Examples</th>
<th># of Items &amp; Responses</th>
<th>Previous Use and Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Refugee Environment Index (PREI)</td>
<td>Assesses a) household resources and b) community social supports available in the camps in Cox’s Bazar. Includes items that focus on caregivers’ perceived support in their immediate environment, especially with respect to child rearing and parenting.</td>
<td>How much do you feel your household has enough income or resources to pay for basic needs like food, clothing, housing?  How much do you feel current accommodation provides enough space for all of the household members?</td>
<td>29 of the original 30 items used 0 “Almost Never” to 3 “Almost Always”</td>
<td>Developed by Pluess and colleagues for the Biological Pathways or Risk and Resilience in the Syrian Refugee Children (BIOPATH) study in Lebanon (McEwan et.al., 2021).</td>
</tr>
<tr>
<td>Caregiver Mental Health</td>
<td>Assesses depressive symptoms.</td>
<td>How often do you feel down, depressed, or hopeless?  How often do you have trouble concentrating on things?</td>
<td>8 of the original 8 items used 0 “Never” to 4 “Almost half the day”</td>
<td>Developed by Spitzer, Williams, and Kroenke (1999) as the PHQ-9 in the United States. The PHQ-8 is the shortened version of the PHQ-9. Validated by Kroenke, et.al. (2009) for the Behavioral Risk Factor Surveillance Survey in the United States. Used by Porter, et. al. (2021) for the Young Lives Study in India, Peru, Vietnam, and Botswana, and in the Rohingya context by Guglielmi, et.al. (2020) for the Gender and Adolescence: Global Evidence (GAGE) longitudinal study. BRAC also uses this in monitoring related to Play to Learn in the Rohingya context.</td>
</tr>
<tr>
<td>Measures</td>
<td>Description</td>
<td>Item Examples</td>
<td># of Items &amp; Responses</td>
<td>Previous Use and Validation</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Parental Stress Scale (PSS)</td>
<td>Assesses how Rohingya parents feel about their children, and experiences of stress specific to child rearing or parenting. Items focussed on parents’ stress such as feelings of being overwhelmed by having children, the economic stress of having children, and concerns related to ability to control one’s own life as a parent.</td>
<td>How much do you feel you have little time/ flexibility since having children?</td>
<td>9 of the original 18 items used 0 “Never” to 4 “Very Often”</td>
<td>Developed by Berry and Jones (1995) in the United States. Further validated by Algarvio, Leal, and Maroco (2018) in Portugal, and Zelmn and Ferro (2018) in Canada.</td>
</tr>
<tr>
<td>Child Behavioral Problem and Competence Scale (CBPCS)</td>
<td>Captures behaviors of children related to competencies and social-emotional problems.</td>
<td>How often does your child make friends easily?  How often does your child act impulsively?</td>
<td>19 items out of 58 Social-Emotional Development Items (EDI) &amp; 42 (BITSEA) 0 “Never” to 4 “Very Often”</td>
<td>This set of items targets a variety of behaviors developed specifically for this study with select items derived primarily from the Early Development Instrument and the BITSEA. The Early Development Instrument was developed by Offord and Janus (2007) in Canada. Validated by Janus, et.al (2007), and further validated by Brinkman, et.al. (2007) in Australia, Ip, et.al. (2013) in China, Duku, Janus, &amp; Brinkman (2015) in Indonesia and the Philippines, and Brinkman, et.al. (2017) in Indonesia BITSE was developed by Briggs-Gowan &amp; Carter (2002), and further validated by Briggs-Gowan and colleagues (2004) in the United States and by Karabekiroglu et.al., (2009) in Turkey.</td>
</tr>
<tr>
<td>Child Positive Behavioral Index (CPBI)</td>
<td>Assesses children’s positive dispositional and interpersonal behaviors.</td>
<td>How often is your child helpful and cooperative?  How often is your child calm, easy-going?</td>
<td>22 of the original 25 items used 0 “Never” to 4 “All the Time”</td>
<td>Developed by Janet Quint, Johannes Bos, and Denise Polit (1997) for the New Chance study in the United States. Further validated by Sylvia Epps, et.al. (2003).</td>
</tr>
</tbody>
</table>
Preparing for the field: Are the measures ready to test?

How do we prepare measures for testing in the field?

Once selected, measures were adapted and translated to be used with the Rohingya population. Adaptation consisted of an extensive process of consultations with local experts who have been working with the Rohingya population since the major influx to Bangladesh in 2017. Through these consultations, we made modifications to the items to make them more relevant and accessible. In some cases, we removed items altogether or chose versions of the scale that were deemed appropriate to the context; in others, the items themselves were modified to be appropriate.

In turn, translation was a multi-step process, given the unique linguistic context of the Rohingya in the context of Cox’s Bazar. English language protocols were first translated into written Standard Bangla and then underwent multiple rounds of revision and training with the enumerators before being implemented in spoken Rohingya. In another research brief in this series, we describe in detail the many considerations involved in the adaptation and translation of materials for the Rohingya context.

What happened during data collection?

We leveraged a study that occurred in collaboration with BRAC (a pilot study related to their Humanitarian Play Labs (HPL) for preschool-aged children) in 2020 and 2021 for data collection. Enumerators from Innovations for Poverty Action (IPA) collected survey data through in-person interviews from 322 Rohingya mothers of 3-4 year old children across 11 camps (no literacy was assumed). This sample was chosen randomly from a larger set of children who attended BRAC’s HPL program. Data collection took place in three phases, which we called ‘mini-pilots’. Changes between mini-pilot one and two focused on language, including trying to confirm conceptual interpretation and understanding of the terminology used in the survey. Changes between mini-pilots two and three focused on enumerator feedback about feasibility (e.g., the length of the survey, whether certain questions were not being understood or were difficult to deliver, and mothers’ response to participation). Between each of these stages, researchers at NYU met frequently with the IPA team to go over what had happened in the field and brainstorm solutions to problems and changes that needed to be made.

Sample

Our target sample included 322 households from 11 camps with 3 or 4 year old children. Phase-1 (n = 42), Phase-2 (n = 64), and Phase-3 (n = 216). 277 (87%) of participants/households completed all of the scales outlined in this brief, with the remaining 45 participants (13%) being caregivers who did not consent to participate. The sample was 100% female with a mean age of 28.95 (SD = 7.16) years. The majority of the caregivers (93%) reported that they attended madrasa/school in Myanmar. There were, on average, 5.76 (SD = 1.82) individuals per household including adults and children. On average, most of the caregivers reported arriving in Bangladesh 3.61 (SD = 2.92) years ago.
Analyzing data: Are our measures usable?

Once we have data, how do we know that the process worked—that this data is usable for analyses, and that the scales can be used again in the future? We examined each measure’s factor structure, reliability, and validity to answer these questions.

Factor structure. Factor analysis is a statistical procedure that allows us to investigate whether assessed items on a survey cohere in meaningful subsets that then can be used as the basis for multi-item, reliable scales. For many scales, such as the PHQ-8, we expected all items in the scale to cohere into one meaningful set. For others, such as the scales examining child social-emotional development, we were interested in exploring whether the scales split up into meaningful sub-scales, as the items in these scales will often be divided into different domains of behavior when used in other contexts. Each of these questions can be explored using factor analysis, which provides a set of indices that can be used to assess what the underlying structure is for the data. We used factor analysis to determine the final items that were included in each scale.

Reliability. Reliability tells us about whether a scale is able to consistently capture the construct. There are many different kinds of reliability. For instance, test-retest reliability informs us on whether the instrument can be used multiple times and give similar results. Internal reliability tells us about whether the different items in a scale that propose to measure the same construct all produce similar scores, i.e., whether a scale is consistent within itself. Here, as an example, we report on internal reliability by calculating a score called Cronbach’s alpha. A higher alpha score is desirable: it indicates that the items are correlated with one another, which means the different items on the scale are most likely measuring the same thing. While cutoffs for alpha scores differ by discipline, in general above 0.6 might be considered an acceptable score, while an alpha of between 0.8 and 0.9 would be very good. We report reliability in Table 2.

Validity. Validity tells us how well an instrument measures what we want it to measure. For example, the validity of an instrument aiming to measure children’s language skills will show us if the instrument actually measures language skills, as opposed to measuring something else, or failing to measure anything at all. While there are several ways to assess validity, it is common, as one initial step, to test for what is called convergent validity: to ensure that the scales that we would expect to be positively correlated with each other are indeed positively correlated, and the scales that we would expect to be negatively correlated are indeed negatively correlated. We report analyses exploring this form of validity in Figure 1.
Results:
How did the scales perform?

In Table 2 below, we provide each scale/sub-scale’s average and standard deviation, reliability score (indicated by Cronbach’s alpha), and a brief description of how we interpret participants’ responses for these scales.

This table also indicates, where relevant, how the scales split into sub-scales. In the interest of being concise, we do not provide full sets of results for factor analyses. Instead we provide our key takeaways; detailed results of factor analyses are available upon request.

Table 2. Summary of descriptives, reliability, and interpretation of scores for each scale.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
<th>Cronbach’s Alpha (reliability)</th>
<th>Score Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceptions of Refugee Camp Environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Resources</td>
<td>0.87</td>
<td>0.34</td>
<td>281</td>
<td>0.75</td>
<td>Higher values mean more access to basic resources</td>
</tr>
<tr>
<td></td>
<td>0.68</td>
<td>0.32</td>
<td>281</td>
<td></td>
<td>Higher values mean more access to community resources</td>
</tr>
<tr>
<td><strong>Caregiver Mental Health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression Symptomatology (PHQ – 8)</td>
<td>1.71</td>
<td>0.70</td>
<td>278</td>
<td>0.80</td>
<td>Higher scores here show higher frequency of reported depression symptoms</td>
</tr>
<tr>
<td>Parenting Stress Scale (PSS)</td>
<td>0.74</td>
<td>0.33</td>
<td>277</td>
<td>0.76</td>
<td>Higher values indicate the mothers were having difficulty with caregiving for their children including feeling stressed and having less time for respite</td>
</tr>
<tr>
<td><strong>Child Social-Emotional Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social (CBPCS)</td>
<td>0.91</td>
<td>0.39</td>
<td>278</td>
<td>0.58</td>
<td>Higher values indicate that the child interacts well with other children including play</td>
</tr>
<tr>
<td>Empathetic (CBPCS)</td>
<td>1.86</td>
<td>0.72</td>
<td>278</td>
<td>0.76</td>
<td>Higher values indicate the child is aware of social cues and engages with both children and adults</td>
</tr>
<tr>
<td>Anxious - Child Behavioral Problem and Competence Scale (CBPCS)</td>
<td>0.98</td>
<td>0.51</td>
<td>278</td>
<td>0.69</td>
<td>Higher values indicate the child may be anxious and afraid to engage with both adult and children</td>
</tr>
<tr>
<td>Externalizing- Child Behavioral Problem and Competence Scale (CBPCS)</td>
<td>1.06</td>
<td>0.39</td>
<td>278</td>
<td>0.66</td>
<td>Higher values indicate the child may be easily irritated and lacks impulse control</td>
</tr>
<tr>
<td>Interactive (CPBI)</td>
<td>1.27</td>
<td>0.47</td>
<td>278</td>
<td>0.76</td>
<td>Higher values indicate that the child is curious and wants to interact with object and people with curiosity</td>
</tr>
<tr>
<td>Regulatory– Child Positive Behavior Index (CPBI)</td>
<td>.99</td>
<td>0.33</td>
<td>278</td>
<td>0.68</td>
<td>Higher values indicate the child is receptive to commands, aware of social surroundings and obedient</td>
</tr>
<tr>
<td>Autonomous– Child Positive Behavior Index (CPBI)</td>
<td>1.53</td>
<td>0.51</td>
<td>278</td>
<td>0.60</td>
<td>Higher values indicate that the child engages in exploration on their own including self-soothing and exploration</td>
</tr>
</tbody>
</table>
Figure 1 shows a correlation matrix between all of the subscales and scales that we discuss here: it illustrates the relationships between each of these scales. This is what we look at to obtain an initial idea about the convergent validity of the scales. For instance, we would expect constructs such as depression and stress to be negatively correlated with better household resources, and positive aspects of children’s behavior to be positively correlated with each other.

Below, see a description of key findings drawn from this table and the correlation matrix.

Figure 1. Correlation matrix indicating relationships between scales and sub-scales.

Sample size (n = 277)
*p < 0.05, **p < 0.01, *** p < 0.001

Color scale indicates magnitude of correlation and direction
- Blue – positively correlated
- Pink – negatively correlated
What have we learned?  
Key Takeaways About Factor Structure, Reliability and Validity

**Factor Structure:** Table 2 indicates the sub-scales that were revealed through factor analyses. For the scales assessing depression (PHQ-8) and parenting stress (PSS), we expected there to be one ‘factor,’ or one meaningful construct emerging from the items of each scale respectively. This was confirmed by analyses. On the other hand, the measure for perceptions of the environment, the PREI, split into two scales, one for household resources and one for community social support. The measures for child social-emotional development also revealed multiple coherent sub-scales through factor analyses: the Child Behavioral Problem and Competence Scale (CBPCS) split into four factors, which we are labeling ‘Social,’ ‘Empathetic,’ ‘Anxious,’ and ‘Externalizing,’ and the Child Positive Behavior Index (CPBI) split into three factors, ‘Interactive,’ ‘Regulatory,’ and ‘Autonomous.’ Table 2 indicates what kinds of behaviors each of these sub-scales maps onto.

**Reliability:** Of the 11 scales, here 6 displayed of our scales displayed acceptable reliability (0.70 and above), while 4 displayed close to acceptable values (0.60-0.70) (as shown by Cronbach’s alpha). Those with stronger reliability – such as PREI and PHQ-8–have been used extensively in humanitarian contexts, which may be a reason they have better reliability in this context as well. Less-used scales that have not been extensively validated in other studies such as the Child Positive Behavior factors and Child Behavioral Problem and Competence Scale factors display promise as a reliable method of assessing child behavior. For the measures showing lower reliability (e.g. social and autonomous child behaviors), next steps may include reexamining translations to see if they are still applicable, or revising the categories or methods through which items are asked to avoid ceiling (uniformly high) or floor (uniformly low) responses.

**Validity:** Relationships between scales are in the hypothesized direction. Most of the correlations between our scales are in the direction that we hypothesized, which is an initial indication of good convergent validity of the measures (see Figure 1). For example, scores on social and empathetic behavior (from the CBPCS) are positively correlated with more positive child behaviors having to do with how interactive, regulated or autonomous a child is on average (as shown by the factors of the CBPI). Another clear example is that households with higher access to basic and community resources (PREI) are negatively correlated with depression scores, indicating a relationship between how many basic and community resources a household has and the caregiver’s reported levels of depression symptoms.
What have we learned?

Conclusions

Preliminary analyses provide reason for optimism in using these particular scales with the Rohingya. Given the participants’ lack of exposure to these kinds of methods (though this perhaps may be changing as exposure to research increases in the camps), it was possible that we would find participants and ask them survey questions but not end up with usable data. Participants might provide answers that are socially desirable, answer the same thing for all items, or become frustrated with what appear to be repetitive and redundant questions—these issues could easily lead to scales that will not yield useful information. Fortunately, with a few exceptions, the measures showed generally strong internal reliability, and the correlational tests to assess validity showed associations in expected directions between caregiver stress and perceived refugee context and between caregiver mental health and child social-emotional development. Overall, the findings indicate that tools measuring caregiver perceptions of environment and early child social-emotional development can be successfully adapted for use in the Rohingya context. This information gives us and our partners the confidence to proceed with piloting other measures from related domains using similar methodologies, and provides other researchers working in this context with much-needed tools.

Further Steps

This brief demonstrates some of the initial steps we take with scales after collecting pilot data, focusing on factor structure, reliability, and validity of the scales. This is not exhaustive. As a first step, we always look at basic descriptive statistics of the scales and items we are interested in—what the average score is, what the variation across participants (i.e. the standard deviation) looks like, and how many participants completed the scale out of our total sample. These basics alert us to initial concerns: a scale, for instance, does not help us much if all participants are scoring exactly the same on it (i.e. there is no variation).

After these basic analyses, we would take further steps before using any of the measures for large-scale data collection. We might: review how variation and range of these scales compare with other researchers’ work, conduct further descriptive analyses to investigate assumptions around whether our pilot sample reflects the intended study sample, examine if results differ by demographics, (e.g., perhaps mothers who have many children answer differently than mothers who have only one) and closely examine patterns of missing data.

Similarly, while a correlation matrix is a first step to show possible relationships between scales, there are also many different statistical modeling techniques that can model relationships and correlations between scales. The ways in which we choose to model these underlying constructs will also have implications for how we frame our hypotheses of child behavior, their environment and interaction with caregivers.
1 **Play to Learn initiative** is an innovative program that harnesses the power of play to deliver critical early learning opportunities to children and caregivers affected by conflict and displacement.


Play to Learn is an innovative program from the LEGO Foundation, Sesame Workshop, BRAC, the International Rescue Committee, and NYU Global TIES for Children that harnesses the power of play to deliver critical early learning opportunities to children and caregivers affected by conflict and crisis. Play to Learn is reaching families affected by the Rohingya and Syrian refugee crises through educational media and direct services in homes, play spaces, health centers, and more to provide the essential building blocks of play based learning and nurturing care. Ultimately, Play to Learn aims to establish play-based early childhood development as an essential component of humanitarian response for all children and caregivers affected by crisis.