



Published in final edited form as:

*SSM Ment Health*. 2022 December ; 2: . doi:10.1016/j.ssmmh.2022.100168.

## Challenges in simultaneous validation of mental health screening tools in multiple languages: Adolescent assessments in Hausa and Pidgin in Nigeria

Bonnie N. Kaiser<sup>a,b,\*</sup>, Cynthia Ticao<sup>c</sup>, Chukwuemeka Anoje<sup>d</sup>, Jeremy Boglosa<sup>c</sup>, Temitope Gafaar<sup>e</sup>, John Minto<sup>c</sup>, Brandon A. Kohrt<sup>b,f</sup>

<sup>a</sup>University of California San Diego, United States

<sup>b</sup>Duke Global Health Institute, United States

<sup>c</sup>Gede Foundation, Nigeria

<sup>d</sup>Catholic Relief Services, Nigeria

<sup>e</sup>Duke University, United States

<sup>f</sup>George Washington University, United States

### Abstract

**Background:** With growing global recognition of the need to address mental health, a key challenge is determining who needs mental health services. Most self-report screening tools were developed in English-speaking high-income settings, and this cultural milieu influences the types and content of items, the manner in which items are asked, and the options for responding to items. Approaches have been developed for transcultural translation and validation. However, these approaches are typically applied in one language at a time, which is of limited utility in linguistically diverse settings.

**Methods:** To address challenges in cross-cultural validation, we undertook a unique process of simultaneously validating tools in two languages in Nigeria. Through this dual-language validation, we explored how cultural and contextual differences may influence what is considered valid for a mental health tool. We validated the Depression Self Rating Scale, Child PTSD Symptom Scale (CPSS), and Disruptive Behavior Disorders Rating Scale with a community sample of 330 adolescents aged 12–17. Validity was assessed in Hausa and Pidgin, two languages commonly spoken in Nigeria. Clinical psychologists used the Kiddie-Schedule for Affective Disorders and Schizophrenia to establish caseness.

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author. 9500 Gilman Road #0532, La Jolla, CA, 92093, United States. [bnkaiser@ucsd.edu](mailto:bnkaiser@ucsd.edu) (B.N. Kaiser).

Author contributions

BNK, BAK, and CT designed the study. CT and JB oversaw data collection, with support from JM, CA, and TG. BNK performed analyses and drafted the manuscript. BAK revised the manuscript. All authors provided feedback.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary data: Appendices A and B

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssmmh.2022.100168>.

**Results:** Most items had good discriminant validity, except on the CPSS, on which only 8 of 17 items discriminated by caseness. Findings indicate the influence of culture (e.g., linguistic differences in translatability of items) and context (e.g., items that reflect experiences of hunger or foodborne illness; different PTSD caseness by language might reflect differential trauma exposure between populations). We also identified items that operated differently between languages.

**Conclusion:** We identified shortcomings in cross-cultural validation procedures with regard to determining whether language, context, or other differences influence performance of items. For future validation efforts, we recommend systematically collecting information on context and stressful/traumatic exposures as a way to contextualize interpretation of the validity findings.

**Acronyms:** Depression Self Rating Scale (DSRS), Child PTSD Symptom Scale (CPSS), Disruptive Behavior Disorders Rating Scale (DBDRS), Oppositional Defiant Disorder (ODD), Conduct Disorder (CD), Area Under the Curve (AUC), Diagnostic Odds Ratio (DOR), Low- and Middle-Income Countries (LMICs), Posttraumatic Stress Disorder (PTSD).

### Keywords

Screening tools Assessment; Validation; Nigeria; Depression; PTSD; Behavioral disorder; Adolescent mental health

---

## 1. Introduction

As low- and middle-income countries (LMICs) continue to expand mental healthcare availability, screening tools are an essential resource for identifying who is in need of care. Particularly in humanitarian contexts, screening tools are an important time-saving resource (Lai et al., 2016). Typically, screening tools are intended for detection and referral efforts, to be followed up with diagnostic evaluation. However, due to the extremely limited number of mental health specialists in many LMIC settings, screening tools have become de facto diagnostic tools (Reynolds and Patel, 2017). This makes it even more essential to ensure validity of such tools. Generally, mental health screening tools have good sensitivity (identifying most or all of those in need of care) but lower specificity (categorizing many non-cases as being in need of care; Mitchell and Coyne, 2007). This reduces efficient and targeted use of resources and can reduce apparent effectiveness of interventions (Kohrt and Kaiser, 2021).

### 1.1. Cultural adaptation and validation

There is increasing recognition of the importance of cultural adaptation as part of the screening tool validation process. Systematic reviews of studies in LMICs found that tools that were locally adapted performed better in validation studies than those that were not adapted (Ali et al., 2016) and that brief screening tools are as accurate as long ones (Akena et al., 2012). Processes for cultural adaptation extend beyond simple translation/back-translation procedures to include annotated translations from multiple lay and professional individuals, qualitative research such as focus group discussions within the target population, and cognitive interviewing procedures (Kaiser et al., 2013; van Ommeren et al., 1999). Some studies also include ethnographic or qualitative methods to identify meaningful mental illness symptoms that are missed by existing screening tools (Kaiser et

al., 2013; Kohrt et al., 2016; Weaver & Kaiser, 2015). The “gold standard” for validating screening tools is clinician diagnosis, although alternative approaches have been tested, such as key informant judgment regarding distress or using interviews by non-clinicians to facilitate evaluation by clinicians at a distance (Bolton, 2001; Watson et al., 2019).

One of the major challenges that is yet to be addressed is how to approach adapting and validating screening tools in places that have wide ethnic and linguistic diversity like Nigeria. Globally, when efforts are made to validate mental health screening instruments, it is typically in one language or is conducted independently in multiple languages that might be spoken in a single setting (Ali et al., 2016). For example, in Nigeria, past research to validate and test screening tools has either been conducted in English or has focused on Yoruba-speaking populations in the southwest of the country (e.g., Adewuya et al., 2007; Omigbodun et al., 1996). Focus on a single language within a linguistically diverse setting undermines efforts to achieve equity in access to mental health-care. At the same time, validating screening tools in multiple languages *independently* raises the risk that screening tools – and their resulting insight regarding referral needs or program effectiveness – might produce different results in different languages within the same setting.

Simultaneous adaptation and validation of screening tools in multiple languages addresses these challenges. However, there is a need to establish procedures for dual-language validation (Kaiser et al. 2019). This raises important questions regarding what validation accomplishes – and how it will subsequently be affected by considering multiple ethnic and linguistic sub-populations simultaneously. Researchers often frame adaptation and validation procedures as focused on capturing *culture*: the language, meanings, or experiences in relation to mental health that are specific to a cultural group (Bolton, 2001; Weaver & Kaiser, 2015). This is important because the terms used to express an experience differ between groups, and the same term may even have different interpretations across groups. Thus, culture shapes the meaning ascribed to particular terms, which influences their endorsement and association with psychological distress or mental illness. For example, cross-cultural research has found that somatic symptoms are salient to mental health in many global settings, i.e., the cultural significance of particular physical complaints is associated with psychological distress or mental illness (Keys et al., 2012; Kleinman, 1977; Ryder et al., 2008; Simon et al., 1999).

At the same time, the specific adaptations that are made often reflect *context* as much as culture. Although there is a wide range of definitions of context, for our purpose we focus on the ecological and structural conditions of available resources and exposure to stressors – for example, the context of a refugee camp or of food insecurity. Culture will influence how forced displacement, war, and famine are experienced, but these material realities cannot be attributed to ‘culture.’ In relation to screening tools, it is common to find that items like stomach aches have poor discriminant validity, which is typically interpreted to reflect the non-specificity of such items in contexts with a high burden of parasites and gastrointestinal infections (Kaiser et al., 2019; Kohrt et al., 2011; Watson et al., 2019). This is because in this context, gastrointestinal distress is common throughout the population and therefore does discriminate between those who do and do not have a mental illness. Thus,

culture and/or context can lead to differences in the significance of somatic complaints for discriminating experience of psychological distress or mental illness.

Consideration of what adaptation and validation procedures accomplish become particularly relevant when conducting dual-language validation. We cannot assume that language is the only relevant variable; sub-populations that differ in primary language might also differ in levels of poverty, cultural history, cultural concepts of distress, trauma exposures, or other factors. In some ways, dual-language validation raises challenges in regards to handling such across-group differences. At the same time, dual-language validation presents an opportunity: to begin to tease apart which differences arising from adaptation and validation procedures reflect linguistic and cultural differences versus contextual differences.

## 1.2. Mental health in Nigeria

Nigeria is the largest African country by population (over 180 million), with half of the population under the age of 25. There are an estimated 17.5 million orphans and vulnerable children in Nigeria, many due to HIV/AIDS (Uneze, 2010). There is widespread recognition that there are enduring psychological effects of AIDS-orphanhood, such as greater rates of depression, anxiety, and post-traumatic stress disorder (PTSD) when compared with other orphans and non-orphans (Cluver et al., 2012; Doku, 2012). On top of a general environment of precarity and insecurity has been added the violence, turmoil, and instability caused by the Boko Haram insurgency. Responsible for tens of thousands killed and millions displaced since 2009, Boko Haram was at one point considered the world's deadliest terror organization (IEP, 2015). In part due to the impact of these events, there has been an increase in adverse childhood outcomes, unsurprisingly including poor mental health outcomes (Atilola, 2012; Omigbodun et al., 2008).

Although population-level estimates do not exist regarding adolescent mental disorders in Nigeria, existing data suggest a high level of treatment need. Studies of children and adolescents in Nigeria using non-representative samples found that 15–20% of children had a current psychiatric disorder, with the majority of these being emotional or conduct disorders (Abiodun, 1992; Omigbodun et al., 1996). These rates are higher than those seen among adults in Nigeria (Adewuya et al., 2018; Gureje et al., 2006). Studies from other sub-Saharan African countries yield similar prevalence estimates of child mental health disorders (Cortina et al., 2012). Significantly, a study of school-based adolescents found that 23% experienced suicidal ideation in the past year, and 12% had attempted suicide – rates among the highest reported in any country (Omigbodun et al. 2008). Despite the greater need among children and adolescents, this population tends to receive less attention in terms of mental health policies, programs, and efforts to validate screening tools for use in the general population and in the midst of humanitarian emergencies.

Currently, there is a shortage of effective mental health assessment tools for use in Nigeria. In particular, there remains a significant need to validate adolescent assessments. Across LMICs, children and adolescents are underrepresented in validation studies for common mental disorder screeners, and there are scarce validation studies for PTSD screeners (e.g., Murray et al., 2011; Ventevogel et al., 2014).

### 1.3. Aims

This study describes a process for validating screening tools in multiple languages simultaneously, which is a significant gap in global mental health literature. In order to have the largest public health impact, we focus on Hausa and Pidgin, two languages commonly spoken by adolescents affected by the Boko Haram crisis. Our primary aim is to simultaneously validate screening tools across languages and produce tools that are brief, easy-to-use, and can be implemented by community-level stakeholders such as community health extension workers. A secondary aim was to explore how cultural and contextual factors influence validation processes, in order to inform broader research methods for cross-cultural validation. The overall goal is to support current and future interventions linking adolescents to mental healthcare, and ultimately to reduce the impact of mental health conditions among vulnerable adolescents. In this study, we focus on depression, posttraumatic stress disorder (PTSD), and behavioral disorders, which are expected to be the most prevalent and burdensome disorders among children and adolescents in Nigeria, particularly those most affected by Boko Haram.

## 2. Methods

### 2.1. Setting

Abuja, the capital of Nigeria, is a planned city intentionally situated at the central point of many ethnic and religious groups. We conducted this study in the linguistically diverse Federal Capital Territory with the aim of producing tools that can be used in various parts of the country. Among the dozens of ethnic groups in Nigeria, Hausa is the largest, constituting approximately 25% of the population. Hausa are largely concentrated in the north, where Boko Haram's impact is greatest (Agbibo, 2014). In addition, there are large populations of Hausa-speaking communities in Abuja who were displaced by the violence between the government and Boko Haram in the northeast (Adewale, 2016). English is the national language, though more often individuals speak a West African Pidgin that combines English terms and grammar with terms and grammar from local West African languages. Hausa is spoken by approximately 48 million people in Nigeria and Pidgin by 30 million, making them two of the most common languages in Africa's largest country (Simons and Fennig, 2018). Therefore, by selecting Hausa and Pidgin as the languages of focus for this study, we anticipate the widest potential public health impact. While most Hausa also speak Pidgin because it is used to communicate across groups, there was minimal overlap between the Hausa speaking communities/respondents and the West African Pidgin speaking communities in our study.

Despite recent increased interest in mental health within Nigeria's policy arena, mental health services remain limited. Recourse to them is typically delayed, often following multiple attempts of care-seeking to community-based providers such as traditional healers (Abdulmalik & Sale, 2012; Abiodun, 1995; Gureje et al., 1995; Agara & Makanjuola, 2006). In line with World Health Organization (WHO) recommendations, Nigeria's 2013 *National Mental Health Policy* recommends task-shifting mental healthcare to non-specialist providers in primary care settings. The success of these programs relies on patients presenting to primary care settings, which can be limited by factors like stigma (Gureje

et al., 2015). Therefore, scholars advocate engaging with local and community-based stakeholders to facilitate linkage of vulnerable individuals to needed mental health services (Abdulmalik et al., 2016; Iheanacho et al., 2015).

This project was conducted by the Gede Foundation and was embedded within the Sustainable Mechanisms for Improving Livelihoods and Household Empowerment (SMILE) program. The SMILE consortium is a cooperative agreement between Catholic Relief Services and the U.S. Agency for International Development, designed to scale-up care and support services for orphans and vulnerable children in four Nigerian states plus the Federal Capital Territory. The program's primary focus areas are household economic strengthening, nutrition, and HIV services. The project described here represents the first step toward incorporating a community-based mental health component into the SMILE program. This study was conducted in nine selected communities from three Area Councils in the Federal Capital Territory where the SMILE program is implemented. These represent particularly vulnerable communities selected for implementation of the SMILE program due to a high burden of socio-economic and health problems, including poverty, malnutrition, and HIV infection.

## 2.2. Instruments

All scales were originally developed in North America or the UK. The Depression Self Rating Scale (DSRS) is an 18-item self-report measure for children and adolescents (Birlleson, 1981). The Child PTSD Symptom Scale (CPSS) was developed as a child version of the Posttraumatic Diagnostic Scale (Foa et al., 1997). The CPSS has 17 items that correspond to PTSD diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders (DSM IV; American Psychiatric Association, 2000). The Disruptive Behavior Disorders Rating Scale (DBDRS) is a 45-item measure corresponding to DSM diagnostic criteria for attention-deficit/hyperactivity disorder (ADHD), oppositional-defiant disorder (ODD), and conduct disorder (CD) (Pelham et al., 1992). Due to the length of the measure, we removed the subscales for inattention and hyperactivity-impulsivity, retaining items relevant to oppositional defiance and conduct disorder. Additionally, the DBDRS was designed to be proxy-administered (by parent or teacher), so we adapted it to be a self-report measure.

In a previous study (Kaiser et al., 2019), we applied an established, systematic process for culturally adapting the selected mental health screening instruments (van Ommeren et al., 1999) that had been modified for use with children and adolescents (Kohrt et al., 2011). We elicited a series of initial translations by trilingual lay and professional individuals. We then conducted focus group discussions (FGDs,  $n = 24$ ) with adolescents to discuss each screening tool item individually. At each stage, translated items were considered in terms of comprehensibility, acceptability, relevance, and completeness. This process aimed to assess equivalence of adapted items to the original English version, in terms of semantic, content, technical, criterion, and conceptual equivalence (van Ommeren et al., 1999).

First, each instrument was translated into Hausa and Pidgin by local researchers not affiliated with this project. In addition to translations, they commented on equivalence of each item (comprehensibility, etc.). Each translation was then reviewed by a team of

four local psychologists, who suggested alternate translations and likewise commented on equivalence. Items were then back-translated to check for completeness, and feedback from translators was used to improve each item. Finally, items were discussed by adolescents in FGDs, stratified by gender, age (12–14/15–17), and language (Hausa/Pidgin). Each FGD reviewed items from one assessment instrument in one language (Hausa/Pidgin). Rather than comparing items to the original English wording, these FGDs asked adolescents to describe each item and to comment on comprehensibility, acceptability, and relevance.

After completion of FGDs, items were again back-translated, reviewed, and further adjusted for clarity. Adapted versions of the screening tools were piloted ( $n = 25$ ) in both Hausa and Pidgin with male and female adolescents between 12 and 17 years. Cognitive interviewing was used, in which participants were asked to respond to each item and then describe their decision-making process ('Why did you give that response?', 'How did you understand that question?'). The purpose of the cognitive interviewing task was to have participants verbalize their interpretation of items in order to identify any items that seemed to be interpreted differently than intended. Following pilot testing, additional adjustments were made to items as needed to improve comprehensibility.

The Translation Monitoring Form (van Ommeren et al., 1999) was used to track adjustments to item translations at each stage of data collection. FGD transcripts were reviewed to identify explicitly stated problems with items, as well as implied problems (e.g., participant giving an example that did not match the item's intended meaning, suggesting that it was not well understood). Notes regarding potential problems with items and how they were addressed were incorporated into the Translation Monitoring Form. Finally, the Form included notes regarding assessments of equivalence at each stage.

Each item was closely examined using the Translation Monitoring Form in both Hausa and Pidgin. Items were adjusted to address any challenges raised throughout data collection, with emphasis on comprehensibility and incorporating specific language suggested by FGD participants wherever possible. Efforts were made to keep items as similar as possible in Hausa and Pidgin. The main reasons for adaptation were items that were conceptually difficult for adolescents to understand, conceptually non-equivalent across languages, considered unacceptable to discuss, or stigmatizing (Kaiser et al., 2019).

### 2.3. Data collection

The culturally adapted screening tools were validated in a community sample of 330 adolescents, aged 12–17, between January–August 2017. Adolescents were recruited by community volunteers, who were instructed to identify adolescents who were either (a) likely experiencing mental, emotional, or behavioral disorder or (b) were mentally healthy. This is similar to processes used in global mental health validation studies among adults and youth (Betancourt et al., 2009; Watson et al., 2019). In order to identify adolescents likely experiencing mental ill-health, community volunteers were given a list of criteria that roughly matched broad categories in the screening tools (e.g., no energy, easily angered, bullying), as well as risk factors (e.g., getting into fights, academic disruptions) that have been predictive of depression among adolescents in other settings in Nigeria (Brathwaite et al., 2020).

The adolescents first completed screening tools with a trained, trilingual (Hausa, Pidgin, English) research assistant. Participants verbally completed each of the 3 screening tools in their choice of language (see Appendix A for screening tool versions following cultural adaptation). Research assistants had undergone two weeks of training in project objectives, quantitative methods, survey data collection, and research ethics. The training included in-depth discussion of each item on the screening tools, supervised practice completing the screening tools, and independent practice with debriefing and feedback. This ensured that all enumerators understood the purpose and process for completing screening tools (e.g., asking items exactly as written) and that they all delivered tools in the exact same way.

Participants were then assessed by one of three Nigerian clinical psychologists, who were blind to results of the screening tools. Clinicians were fluent in Hausa and Pidgin. Clinical assessment was based on the Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) (Kaufman et al., 1997). The K-SADS is a child and adolescent version of the adult Schedule for Affective Disorders and Schizophrenia (Endicott and Spitzer, 1978). It is a semi-structured diagnostic interview that allows trained interviewers to score children and adolescents on DSM diagnoses. The K-SADS modules administered in the study included depression, PTSD, and disruptive behavior disorders (CD and ODD). In addition, the K-SADS is used to collect information on impairment in personal, academic, family, peer, and occupational functioning.

The Nigerian clinical psychologists involved in this study were trained for 3 days on the K-SADS by the senior author (BAK), who is a psychiatrist with experience in training on the K-SADS and other structured clinical interviews in diverse cultural populations. The training included reviewing each module and symptom probe. The clinicians observed K-SADS administrations. Then, they interviewed adolescents with the different disorders to practice using the tool. Finally, inter-rater agreement was assessed with another group of adolescents, each of whom was interviewed by two clinicians in separate consecutive interviews, during which the clinicians were blind to their colleagues' assessment. Each clinician interviewed 3 patients to allow for pairwise comparisons among all clinicians. Inter-rater agreement was assessed on each symptom of the included modules. The clinicians reached 98% inter-rater agreement before independently assessing adolescents. Screening and clinical interviews together took approximately 30–45 min per participant.

**2.3.1. Caseness criteria**—Clinical interviews were used to establish a gold standard comparison with which to validate the screening tools, by identifying those meeting clinical criteria for each disorder (depression, PTSD, ODD, CD). The K-SADS modules can be used to indicate caseness (indicating that an adolescent likely has a disorder) as well as meeting full diagnostic criteria. Caseness was used rather than diagnosis because the purpose of most screening tools is to identify individuals with likely disorder and who therefore require more clinical assessment. In order to match the validation criteria to the purpose of the tool (i.e., optimizing sensitivity), we used caseness rather than full diagnostic criteria in validating the tools.

In K-SADS modules, caseness is determined by functional impairment and meeting criteria for minimum number of symptoms at subthreshold levels. Full diagnosis is based on



functional impairment and meeting symptoms criteria on a number of symptoms at threshold levels. Each symptom assessment includes a description of subthreshold and threshold. For example, the “depressed mood” subthreshold criterion is “often experiences dysphoric mood at least 3 times a week for more than 3 h each time”; the threshold criterion is “feels “depressed” most of the day more days than not.”

Depression caseness requires functional impairment, plus subthreshold levels for 2 weeks on at least one of the following Group A symptoms: depressed mood; irritability or anger; anhedonia, lack of interest, low motivation, or boredom; recurrent thoughts of death; suicidal ideation; suicidal acts; or non-suicidal physically damaging acts; and 3 symptoms from among Group B symptoms: sleep disturbances; fatigue, lack of energy, tiredness; cognitive disturbances; appetite/weight changes; psychomotor agitation/retardation; negative self-perceptions; and hopeless, helplessness, discouragement, and pessimism. Depression diagnosis requires functional impairment plus threshold levels for at least one Group A symptom and 5 Group B symptoms.

PTSD caseness requires experiencing a traumatic event plus functional impairment plus endorsing one of the following for at least one month: recurrent thoughts or images of event; efforts to avoid thoughts or feelings associated with the trauma; nightmares; insomnia; irritability or outbursts of anger. PTSD diagnosis requires experiencing a traumatic event plus functional impairment plus endorsing at least one of the caseness items above plus the following for at least one month: at least one of the re-experience items; at least three of the persistent avoidance items; and at least two of the increased arousal items.

ODD caseness was operationalized as functional impairment plus at least one subthreshold level symptom among Group A symptoms: loses temper, argues a lot with adults, disobeys rules a lot, plus subthreshold levels on three of the following Group B symptoms for at least 6 months: easily annoyed or angered, angry or resentful, spiteful and vindictive, uses bad language, annoys people on purpose, and blames others for own mistakes. Diagnosis of ODD requires one threshold level Group A symptom and three Group B symptoms.

CD caseness was operationalized as functional impairment plus at least one subthreshold level symptom among Group A symptoms: lies; truant; initiates physical fights; bullies, threatens, or intimidates others; or nonaggressive stealing, plus subthreshold levels on three of the following Group B symptoms for at least 6 months: vandalism; breaking and entering; aggressive stealing; often stays out at night after curfew; ran away overnight; use of weapon; physical cruelty to persons; forced sexual activity; or cruelty to animals. Diagnosis of CD requires one threshold level Group A symptom and three Group B symptoms.

#### 2.4. Data analysis

Screening tool data were analyzed for descriptive statistics, internal consistency using Cronbach’s alpha, and item-total correlations. Independent samples t-tests were performed to compare group means between Pidgin and Hausa versions of the overall screening tools, as well as between cases and non-cases (established according to clinical interviews). To assess discriminant validity, case/non-case t-tests used a cut-off of  $p < 0.20$ . We selected a more liberal cut-off because we wanted to retain as many items from the original scales as

possible. Caseness analyses were based on the overall dataset (Pidgin and Hausa combined) because there were not enough cases for either language to examine caseness discrimination for each language separately.

The between-language t-tests were used to identify significant differences in item scores between languages; such differences would not be expected if items perform comparably between languages. For diagnoses where case proportion differed by language group (e.g., more participants classified as depression cases among Hausa compared to Pidgin participants), items on the corresponding screener were considered problematic if item endorsement was in the opposite direction as caseness (e.g., if scores on a DSRS item were higher among Pidgin than Hausa participants). For analysis, DBDRS items were split into those corresponding to ODD and those corresponding to CD, which is a standard approach for scoring the tool. For items that either did not discriminate by K-SADS caseness or differed by language in the opposite direction as overall caseness, they were removed from the screening tool. Sum scores on each screener were calculated at the level of the individual. Missingness was minimal, with 2 or 3 total item-responses missing per screener among the whole sample.

Potential cut-off scores were assessed in comparison to clinical caseness to calculate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive/negative likelihood odds ratios (LR+/LR-), diagnostic odds ratio (DOR), area under the Receiver Operating Characteristic curve (AUC-ROC), and Youden's index (YJ). Sensitivity refers to a screening tool's ability to detect true cases (i.e. children who are categorized as cases by a clinician) and is expressed as a percentage out of 100. Specificity refers to a screening tool correctly identifying true non-cases. Positive predictive value refers to the percentage of positive screens that are true cases, while negative predictive value is the percentage of negative screens that are true non-cases. A positive likelihood ratio is a measure of how many times more likely a positive screen is within a case than a non-case, while a negative likelihood ratio indicates how many times less likely a negative screen is within a case than a non-case. A diagnostic odds ratio is the ratio of the odds of a positive screen among cases to the odds of a positive screen among non-cases. AUC-ROC indicates how well a diagnostic test classifies true cases relative to false positives. Youden's index is a global measure of screening test performance that combines sensitivity and specificity. Together, these measures inform selection of the ideal cut-off score for each screening tool, or the minimum total score that indicates a need for referral to services (Šimundi , 2009). While we considered all psychometric properties holistically, our aim in selecting cut-off scores was to maximize specificity while maintaining a sensitivity >80%. Final versions of the screeners are available in Appendix B.

## 2.5. Ethical considerations

All study procedures were approved by the Federal Capital Territory Health Research Ethics Committee (Approval Number FHREC, 2016/01/44/22/06/16). Before data collection, community chiefs provided *loco parentis* consent on behalf of adolescents, which is a common and accepted approach in Nigeria. All adolescents assented before participating.

### 3. Results

#### 3.1. Sample

Participants were 75% male, with a mean age of 15 (Table 1). More adolescents were assessed in Hausa ( $n = 194$ ) than in Pidgin ( $n = 136$ ). Most lived with their parents, and slightly over half had attended up to a primary education (equivalent of elementary school). Hausa respondents reported larger mean family size (9 vs 7) and were more likely to have secondary education (48% vs 35%), while Pidgin respondents were more likely to be living with both parents (79% vs 68%). In the PTSD screener, the most common distressing event named by participants was the death of a family member or friend. These events were proportionally similar by language of participant (though they were only asked to name their most distressing event, rather than being asked if they experienced each event). Fewer than 3% of participants reported *never* having experienced a distressing event.

#### 3.2. Depression Self Rating Scale (DSRS)

Among adolescents assessed in Hausa ( $n = 194$ ), 15 (7.8%) were categorized as clinical cases, compared with 6 (4.4%) of those assessed in Pidgin ( $n = 136$ ); this difference was not statistically significant. Of the original 18 items, 4 did not discriminate between cases and non-cases (Table 2). Three of these items also had low or negative item-total correlations, after correcting for reverse scoring. Seven items differed significantly by language, but these differences were almost always in the expected direction based on case proportion of each language. Ultimately, 4 items were removed from the final version of the screening tool based on lack of discrimination of mental health cases and non-cases. The resulting Cronbach's  $\alpha = 0.77$  and area under the curve (AUC) = 0.71, which is the same as the AUC for the 18-item version.

#### 3.3. Child PTSD Symptom Scale (CPSS)

Among adolescents assessed in Hausa, 30 (15.5%) were categorized as clinical cases, compared with 11 (8.1%) of those assessed in Pidgin ( $p < 0.05$ ). Of the original 17 items, 9 did not distinguish between cases and non-cases (Table 3). One of these items also had a negative item-total correlation, and endorsement differed significantly by language in the opposite direction of expected based on caseness proportion. All of these items were removed from the final version of the screening tool. The resulting  $\alpha = 0.73$  and AUC = 0.67, compared to an AUC = 0.63 for the 17-item version.

#### 3.4. Disruptive Behavior Disorders Rating Scale (DBDRS)

Among adolescents assessed in Hausa, 33 (17.2%) were categorized as clinical ODD cases, compared with 25 (18.4%) of those assessed in Pidgin. For CD, 61 (31.8%) Hausa speakers were classified as cases, as were 30 (22.1%) of those assessed in Pidgin. Of the original 8 ODD items, 1 did not discriminate between cases and non-cases overall (Pidgin and Hausa combined; Table 4). It also had low or negative item-total correlations. This item was removed from the final version of the screening tool. The resulting  $\alpha = 0.69$  and AUC = 0.70, compared to an AUC = 0.68 for the 8-item version. Of the original 15 CD items, 4 either did not discriminate between cases and non-cases or differed significantly by language

in the opposite direction of expected based on differences in case proportion (Hausa 32% vs Pidgin 22%,  $p = 0.05$ ; Table 5). One of these items also had a low item-total correlation. All of these items were removed from the final version of the screening tool. The resulting  $\alpha = 0.82$  and  $AUC = 0.76$ , compared to an  $AUC = 0.77$  for the 15-item version.

### 3.5. Selecting cut-off scores

Table 6 provides the psychometric properties based on cut-off scores selected for each screening tool. The sensitivity values are sufficient to recommend use of these screening tools to detect mental disorder caseness among adolescents in Nigeria. Specificity scores are low; however, as screening tools are not intended to be diagnostic but to identify individuals in need of additional evaluation, we prioritized high sensitivity despite lower specificity.

## 4. Discussion

The lack of culturally valid mental health assessment instruments is a major barrier to screening individuals into mental health interventions and evaluating their effectiveness. We went beyond typical approaches to adaptation and validation by validating screening tools in parallel in multiple languages. Our pragmatic multi-linguistic approach identifies items that function equivalently across language, as well as highlighting items that perform differently across languages.

### 4.1. Discriminant validity

Overall, most items on all screeners performed well in terms of discriminant validity, except on the CPSS, where only 8 of the 17 items were discriminant. Only the DBDRS subscales had high enough numbers of cases for each language group to assess discriminant validity separately by language. For both the ODD and CD subscales, about half of the items had similar discriminant validity for both languages, while the other half were discriminant in only one language or the other. Similarities and differences in discriminant validity of items allow us to consider whether validation findings reflect culture or context, including factors like trauma exposure, all of which might differ by sub-population.

Items that discriminated caseness align with existing literature. For example, one of the best discriminating items used an idiom of distress (CPSS-11: “having a dry heart”). Scholars advocate use of idioms of distress to improve comprehensibility and validity of screening tools (Kaiser et al., 2013; Kohrt et al., 2016). Three of the best discriminating DSRS items were crying, loneliness, and sadness, which are important in cross-cultural phenomenology of depression (Haroz et al., 2017). Loneliness was also highlighted as an important culturally salient indicator of depression among adolescents in Lagos, Nigeria (Ottman et al., 2022). Items regarding bad dreams functioned well on both the DSRS and CPSS, while trouble sleeping did not discriminate on either screener (Hinton and Lewis-Fernández, 2011). Finally, items about food (DSRS-6 “tummy aches” and DSRS-8 “enjoy food”) were discriminant after adjusting items to exclude experiences associated with widespread foodborne illness and food insecurity – changes that have likewise been required in other settings (Kohrt et al., 2011).

Other items were conceptually complex, and changes required to improve comprehensibility might have explained their poor discriminant functioning. For example, translating abstract concepts led some items to become relatively long (e.g., CPSS-5). Additionally, item valence seemed to affect functioning. On the DSRS, all non-discriminant items were positively worded, and the one CPSS item that was reworded as positive during the adaptation process was ultimately non-discriminant. Other studies have similarly identified problems with positive valence items, which might generate confusion for participants particularly when valence changes within a screener (Kohrt et al., 2016; Watson et al., 2019; Weobong et al., 2009). In contrast, the best functioning items were concrete, as suggested by the fact that the behavioral tool (DBDRS) functioned best overall.

#### 4.2. Culture and context

Some items' performance seems to reflect differential ability to capture cultural or linguistic meaning of mental health experiences. For example, CPSS-11 ("inability to have strong feelings") was asked with an idiom of distress in Hausa ("having a dry heart") but not in Pidgin (Kaiser et al., 2019). This item had discriminant validity in Hausa but not Pidgin (data not reported). This might be due to the low number of cases in Pidgin, or it might suggest that culture – specifically the extent to which cultural adaptation is successful in reflecting local meanings and language – might partly account for differential discriminant validity between linguistic groups. Similarly, CPSS items regarding re-experiencing and avoidance did not discriminate between cases and non-cases. These findings might reflect difficulty in translating meaning, or they might suggest that such experiences are not key aspects of PTSD experiences in this setting, as has been reported elsewhere (McCall and Resick, 2003). Differential discriminant validity regarding DBDRS items might reflect differences in acceptability of certain behaviors within linguistic groups. For example, DBDRS-3 ("disobeys adults") was discriminant only for Pidgin and was also endorsed more highly in Pidgin (whereas most items were endorsed similarly between languages). This might suggest that disobeying adults is more acceptable to endorse in one cultural group than another.

Context likewise helps to explain findings, particularly when they differ between language groups. For example, two items related to food were found to be problematic during the cultural adaptation process and were adjusted to focus on the meaning intended by the original English items, rather than unrelated experiences such as hunger. DSRS-6 (tummy aches) was adjusted to clarify "not caused by hunger or sickness," and DSRS-8 (enjoying food) was clarified by adding "when food is available" (Kaiser et al., 2019). These items were endorsed more highly among Hausa than Pidgin respondents, perhaps due to differences in context regarding food availability, quality, or foodborne illnesses. Alternatively, as with CPSS-11, these differences might reflect differential success in translating items' meaning between Hausa and Pidgin. However, both items were ultimately discriminant by caseness. Additionally, items regarding sleep (DSRS-2 "sleep well" and CPSS-13 "trouble sleeping") did not discriminate caseness. Such findings might reflect that problems sleeping are non-specific to experiences of mental distress in this context. Other items appeared to reflect contextually specific expectations. For example, in qualitative data collection, CPSS-16 ("overly cautious") was interpreted as reflecting an appropriate

degree of caution for this hazardous environment (Kaiser et al., 2019). This item did not discriminate by caseness. Similarly, CPSS-3 (“re-experiencing”) was difficult for adolescents to distinguish from feeling frightened that the same traumatic event might happen again, which might reflect a context of vulnerability and expectations of trauma experiences.

Finally, contextual factors like exposure to stressors and traumatic events might have explanatory value regarding differential discriminant validity by language. The two subgroups differed in terms of caseness for PTSD (Hausa: 15% vs Pidgin: 8%) and conduct disorder (32% vs 22%). This might reflect differences in trauma exposures, although our data cannot speak to that. For example, although our recruitment process did not differ by language, it is possible that we systematically recruited more “problem” adolescents (e.g., with worse trauma exposures) who speak Hausa than is reflected more broadly in the population. This reflects a general shortcoming of cross-cultural validation studies, in that we are unable to assess differential validity by nuanced sub-population characteristics, including trauma exposure as well as factors like gender and developmental stages, which could influence discriminant validity of items. In future, such research should more systematically assess such demographic and developmental factors, as well as stressors and trauma exposures, in order to understand population differences that might factor into validation success.

#### 4.3. Contrasts with existing literature

A significant finding is that rates of behavioral disorders (ODD and CD) in our sample differed from research in other countries and prior validation research among adolescents in Nigeria. Using the K-SADS among primary care patients in southwest Nigeria, Omigbodun et al. (1996) found equal rates of depression and conduct disorder (6%), while 18–28% of our sample was classified as behavioral disorder cases, compared to 6.5% classified as depression cases. These findings might reflect the older age of our sample (12–17 compared to 10–14 in Omigbodun et al.), the use of caseness rather than diagnostic cut-offs in our study, or the effects of disparate social, economic, and political turmoil both over time and regionally in Nigeria. Although our depression caseness findings are on par with Omigbodun et al., it might also be the case that depression symptoms went under-detected in our sample.

Additionally, we found higher rates of CD caseness (28%; behaviors like lying, initiating fights, and stealing) compared to ODD (18%; behaviors like arguing with adults and disobeying rules), which does not match findings in most other settings (Matthys and Lochman, 2017 but see Canino et al. 2010 for exceptions). These findings might reflect the local eco-cultural context, specifically cultural norms regarding adolescents’ behaviors (Burkey et al., 2016; Super and Harkness, 1986). For example, arguing with or disobeying an adult (symptoms of ODD) is considered strongly taboo. In contrast, behaviors like fighting, lying, and stealing – particularly among peers – might represent means of expressing distress that are more socially sanctioned in Nigeria. Additionally, in the study region, there is an increase in general violence and adolescents’ engagement in gangs. In some North American and Western European cultural groups, a lenient family environment

might make symptoms of ODD more culturally acceptable, while well-functioning legal systems generate strong deterrents to CD behaviors. In contrast, settings like Nigeria are marked by strong social sanctions against ODD behaviors within the family, whereas the normalization of violent or criminal behavior combined with a weak legal system might allow CD behavior to be more prominent.

#### 4.4. Limitations

We faced challenges recruiting enough “problem” adolescents, or those considered likely to have a mental health disorder. This is likely because community volunteers primarily recruited participants from the SMILE program, who had to meet requirements of higher functioning for school readiness. Ultimately, we adapted our recruitment criteria to include non-beneficiaries. We had more cases with behavioral than emotional problems, and our validation sample was heavily skewed towards males. This is likely because externalizing problems were easier for volunteers to identify. We did not have a large enough number of cases to run caseness analyses separately by language. One reason that individual items may have not discriminated caseness was because of differences between samples. However, given the small samples, we cannot differentiate whether items were not associated with the conditions of interest or whether items had different relationships with the condition of interest based on the language and population. The significant reduction in items for some of the final screening tools may have affected the construct domain coverage of the final screening tools; however, we attempted to account for this by ensuring that key diagnostic items for each disorder were retained.

As with all efforts to adapt and validate screening tools, there are cultural and contextual considerations, such as participants’ lack of familiarity with Likert-type scales and potential social desirability bias because all tools were verbally administered. One of the challenges in global mental health studies of instrument adaptation and validation is that gold standards for validation (K-SADS in this study) also have cultural biases. Although we trained Nigerian psychologists in the K-SADS and extensively discussed each item and established strong inter-rater reliability, the Nigerian K-SADS itself was not validated against another standard. Another limitation is that the study used the DSM-IV version of the CPSS and comparable K-SADS PTSD criteria. The more complicated and restrictive criteria of DSM5 with 5 domains of symptom types would have likely reduced the prevalence of PTSD compared to what was observed in this study.

Although the diversity of the Federal Capital Territory allowed us to validate screeners in multiple languages, there are likely regional differences that should be addressed if these screeners are used elsewhere. Specifically, although Pidgin is a national language, Pidgin dialects differ somewhat across the country. Efforts to apply the screening tool in other regions should first ensure local comprehensibility and relevance of the Pidgin version. We advocate that similar efforts at simultaneous multi-language validation be made in other places that, like Nigeria, have wide ethnic and linguistic diversity, in order to ensure representation and equity in mental health detection and care.

#### 4.5. Applications and recommendations

We produced linguistically ipsitized screening tools for which sensitivity and specificity are balanced across groups without requiring separate tools. This has been a major gap in areas of ethnic and linguistic diversity like Nigeria and should be pursued more often in global mental health research and practice. In conflict settings, it is necessary to have tools that are predictive across sub-populations, making it infeasible to conduct separate validation studies for sub-populations.

More multi-language translation and validation studies are needed to expand this important literature. In particular, future studies should explore issues of culture, context, and exposures through gathering data regarding factors such as level of education, physical health, and traumatic and stressful exposures in addition to detailed linguistic data. Our study suggests that these factors – particularly differential exposures across linguistic groups – could be important for explaining differences in caseness and discriminant validity. Validation studies that collect data on culture, context, and exposures could shed light on their influence on psychometric properties and validity of assessment tools.

In addition to these quantitative data, studies should make strategic use of qualitative data that can shed light on *why* these factors matter. For example, qualitative research is central to cultural adaptation procedures, and there is a rich literature demonstrating how such procedures improve acceptability and validity of assessment tools (Ali et al., 2016; Kaiser et al., 2019; Kohrt et al., 2016). Such approaches address culture and context, but qualitative research could also provide insight into why and how differential trauma and stress exposures arise between linguistic groups and how they affect validity of assessment tools. Finally, future multi-language validation studies should consider additional quantitative approaches. Methods such as network analysis, item response theory, and measurement invariance could advance understanding of assessment tools' functioning across languages (Borsboom, 2017; Putnick and Bornstein, 2016; Reise and Waller, 2009).

### 5. Conclusion

We evaluated the psychometric properties of transculturally adapted versions of mental health tools for adolescents in central Nigeria. This research is particularly needed as the social, economic, and psychological effects of the Boko Haram insurgency continue. Our study raises questions about the roles of culture, context, and stress and trauma exposures in validation studies. Our findings point toward the need for new approaches in global mental health validation to ensure that tools are best suited to identify who is need of services, especially for resource constrained settings. Our approach also provides guidance for validating standardized tools that require comparable use across linguistically diverse populations.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgements

We thank Drs. Andrew Zamani, Olayinka Omigbodun, and Michael Ezenwa for feedback on an earlier version of this manuscript. We thank all participants in the Stakeholders Meeting for valuable feedback on this study. We are grateful to our research assistants and all those who participated in this study. Support for this paper was provided by Catholic Relief Services Implemented SMILE Project with funds from the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through U.S. Agency for International Development (USAID) Cooperative Agreement No. AID-620-A-13-00003 and technical support from Gede Foundation. The views expressed in this publication do not necessarily reflect those of CRS or USAID. Dr. Kaiser was supported by the National Institute of Mental Health of the National Institutes of Health (F32MH113288). Dr. Kohrt and Dr. Kaiser were supported by the US NIMH (K01MH104310, R21MH111280).

## References

- Abdulmalik J, Kola L, & Gureje O (2016). Mental health system governance in Nigeria: challenges, opportunities and strategies for improvement. *Glob. Ment. Health*, 3.
- Abdulmalik JO, & Sale S (2012). Pathways to psychiatric care for children and adolescents at a tertiary facility in northern Nigeria. *J. Publ. Health Afr*, 3(1).
- Abiodun OA (1995). Pathways to mental health care in Nigeria. *Psychiatr. Serv*, 46(8), 823–826. [PubMed: 7583485]
- Abiodun OA (1992). Emotional illness in a paediatric population in Nigeria. *East Afr. Med. J*, 69, 557–559. [PubMed: 1473508]
- Adewale S (2016). Internally displaced persons and the challenges of survival in Abuja. *Afr. Secur. Rev*, 25(2), 176–192.
- Adewuya A, Ola B, & Aloba O (2007). Prevalence of major depressive disorders and a validation of the Beck Depression Inventory among Nigerian adolescents. *Eur. Child Adolesc. Psychiatr*, 16(5), 287–292.
- Adewuya A, Coker O, Atilola O, Ola B, Zachariah M, Adewumi T, & Idris O (2018). Gender Difference in the Point Prevalence, Symptoms, Comorbidity, and Correlates of Depression: Findings from the Lagos State Mental Health Survey (LSMHS), Nigeria (pp. 1–9). *Archives of Women's Mental Health*.
- Agara AJ, & Makanjuola AB (2006). Pattern and pathway of psychiatric presentation at the out-patient clinic of a neuro-psychiatric hospital in Nigeria. *Niger. J. Psychiatr*, 4(1), 30–34.
- Agbibo DE (2014). Peace at daggers drawn? Boko Haram and the state of emergency in Nigeria. *Stud. Conflict Terrorism*, 37(1), 41–67.
- Akena D, Joska J, Obuku EA, Amos T, Musisi S, & Stein DJ (2012). Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries: a systematic review. *BMC Psychiatr.*, 12(1), 187.
- Ali GC, Ryan G, & De Silva MJ (2016). Validated screening tools for common mental disorders in low and middle income countries: a systematic review. *PLoS One*, 11(6), Article e0156939.
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Washington, DC: APA.
- Atilola O (2012). Can family interventions be a strategy for curtailing delinquency and neglect in Nigeria? Evidence from adolescents in custodial care. *Afri. J. Psychol. Stud. Soc. Iss*, 15, 218–237.
- Betancourt TS, Bass J, Borisova I, Neugebauer R, Spielman L, Onyango G, & Bolton P (2009). Assessing local instrument reliability and validity: a field-based example from northern Uganda. *Soc. Psychiatr. Psychiatr. Epidemiol*, 44(8), 685–692.
- Birleson P (1981). The validity of depressive disorder in childhood and the development of a self-rating scale: a research report. *JCPP (J. Child Psychol. Psychiatry)*, 22(1), 73–88. [PubMed: 7451588]
- Bolton P (2001). Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *J. Nerv. Ment. Dis*, 189(4), 238–242. [PubMed: 11339319]
- Borsboom D (2017). A network theory of mental disorders. *World Psychiatr.*, 16(1), 5–13.

- Brathwaite R, Rocha TBM, Kieling C, Kohrt BA, Mondelli V, Adewuya AO, & Fisher HL (2020). Predicting the risk of future depression among school-attending adolescents in Nigeria using a model developed in Brazil. *Psychiatr. Res*, 294, Article 113511.
- Burkey M, Ghimire L, Adhikari R, Kohrt B, Jordans M, Haroz E, & Wissow L (2016). Development process of an assessment tool for disruptive behavior problems in cross-cultural settings: the Disruptive Behavior International Scale – Nepal version (DBIS-N). *Int. J. Cult. Mental Health*, 9(4), 387–398.
- Canino G, Polanczyk G, Bauermeister J, Rohde L, & Frick P (2010). Does the prevalence of CD and ODD vary across cultures? *Soc. Psychiatr. Epidemiol.*, 45(7), 695–704.
- Cluver L, Orkin M, Gardner F, & Boyes ME (2012). Persisting mental health problems among AIDS-orphaned children in South Africa. *JCPP (J. Child Psychol. Psychiatry)*, 53, 4.
- Cortina MA, Sodha A, Fazel M, & Ramchandani PG (2012). Prevalence of child mental health problems in sub-Saharan Africa: a systematic review. *Arch. Pediatr. Adolesc. Med.*, 166, 276–281. [PubMed: 22393184]
- Doku P (2012). *The Mental Health of Orphans and Vulnerable Children within the Context of HIV/AIDS in Ghana*. Scotland: Doctoral dissertation, University of Glasgow.
- Endicott J, & Spitzer RL (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch. Gen. Psychiatr.*, 35(7), 837–844. [PubMed: 678037]
- Foa EB, Cashman L, Jaycox L, & Perry K (1997). The validation of a self-report measure of posttraumatic stress disorder: the Posttraumatic Diagnostic Scale. *Psychol. Assess.*, 9(4), 445–451.
- Gureje O, Abdulmalik J, Kola L, Musa E, Yasamy MT, & Adebayo K (2015). Integrating mental health into primary care in Nigeria: report of a demonstration project using the mental health gap action programme intervention guide. *BMC Health Services Research*, 15, 242. [PubMed: 26094025]
- Gureje O, Acha RA, & Odejide OA (1995). Pathways to psychiatric care in Ibadan, Nigeria. *Tropical and Geographical Medicine*, 47(3), 125–129. [PubMed: 7483003]
- Gureje O, Lasebikan VO, Kola L, & Makanjuola VA (2006). Lifetime and 12-month prevalence of mental disorders in the Nigerian survey of mental health and well-being. *Br. J. Psychiatry*, 188, 465–471. [PubMed: 16648534]
- Haroz E, Ritchey M, Bass J, Kohrt B, Augustinavicius J, Michalopoulos L, & Bolton P (2017). How is depression experienced around the world? A systematic review of qualitative literature. *Soc. Sci. Med.*, 183, 151–162. [PubMed: 28069271]
- Hinton D, & Lewis-Fernández R (2011). The cross-cultural validity of posttraumatic stress disorder: implications for DSM-5. *Depress. Anxiety*, 28(9), 783–801. [PubMed: 21910185]
- Iheanacho T, Obiefune M, Ezeanolue CO, Ogedegbe G, Nwanyanwu OC, Ehiri JE, ... Ezeanolue EE (2015). Integrating mental health screening into routine community maternal and child health activity: experience from prevention of mother-to-child HIV transmission (PMTCT) trial in Nigeria. *Social Psychiatry and Psychiatric Epidemiology*, 50(3), 489–495. [PubMed: 25201181]
- Institute for Economics and Peace. (2015). *Global Terrorism Index 2015 (PDF)*. November 2015. Retrieved. (Accessed 7 January 2018).
- Kaiser BN, Kohrt B, Keys H, Khoury N, & Brewster A (2013). Strategies for assessing mental health in Haiti: local instrument development and transcultural translation. *Transcult. Psychiatr.*, 15(4), 532–558.
- Kaiser BN, Ticao C, Anoje C, Minto J, Boglosa J, & Kohrt BA (2019). Adapting culturally appropriate mental health screening tools for use among conflict-affected and other vulnerable adolescents in Nigeria. *Glob. Mental Health*, 6.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J. Am. Acad. Child Adolesc. Psychiatry*, 36(7), 980–988. [PubMed: 9204677]
- Keys HM, Kaiser BN, Kohrt BA, Khoury NM, & Brewster ART (2012). Idioms of distress, ethnopsychology, and the clinical encounter in Haiti's Central Plateau. *Soc. Sci. Med.*, 75(3), 555–564. [PubMed: 22595073]

- Kleinman A (1977). Depression, somatization and the “new cross-cultural psychiatry. *Soc. Sci. Med.*, 11, 3–10. [PubMed: 887955]
- Kohrt BA, Jordans M, Wietse A, Nagendra P, Maharjan M, & Upadhaya N (2011). Validation of cross-cultural child mental health and psychosocial research instruments: adapting the Depression Self-Rating Scale and Child PTSD Symptom Scale in Nepal. *BMC Psychiatr*, 11, 127.
- Kohrt BA, & Kaiser BN (2021). Measuring mental health in humanitarian crises: a practitioner’s guide to validity. *Conflict Health*, 15(1), 1–10. [PubMed: 33390172]
- Kohrt B, Luitel N, Acharya P, & Jordans M (2016). Detection of depression in low resource settings: validation of the Patient Health Questionnaire (PHQ-9) and cultural concepts of distress in Nepal. *BMC Psychiatr*, 16, 58.
- Lai B, Alisic E, Lewis R, & Ronan K (2016). Approaches to the assessment of children in the context of disasters. *Curr. Psychiatr. Rep.*, 18(5), 45.
- Matthys W, & Lochman J (2017). *Oppositional Defiant Disorder and Conduct Disorder in Childhood.* John Wiley & Sons.
- McCall GJ, & Resick PA (2003). A pilot study of PTSD symptoms among Kalahari Bushmen. *J. Trauma Stress*, 16, 445–450. [PubMed: 14584628]
- Mitchell AJ, & Coyne JC (2007). Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br. J. Gen. Pract.*, 57(535), 144–151. [PubMed: 17263931]
- Murray L, Bass J, Chomba E, Imasiku M, Thea D, Semrau K, et al. (2011). Validation of the UCLA Child Post traumatic stress disorder-reaction index in Zambia. *Int. J. Ment. Health Syst*, 5(1), 24. [PubMed: 21943178]
- Omigbodun O, Dogra N, Esan O, & Adedokun B (2008). Prevalence and correlates of suicidal behaviour among adolescents in southwest Nigeria. *Int. J. Soc. Psychiatr*, 54(1), 34–46.
- Omigbodun O, Gureje O, Ikuesan B, Gater R, & Adebayo E (1996). Psychiatric morbidity in a Nigerian paediatric primary care service: a comparison of two screening instruments. *Soc. Psychiatr. Psychiatr. Epidemiol*, 31(3–4), 186–193.
- Ottman K, Wahid SS, Flynn R, Momodu O, Fisher HL, Kieling C, Mondelli V, Adewuya A, & Kohrt BA (2022). Defining culturally compelling mental health interventions: a qualitative study of perspectives on adolescent depression in Lagos, Nigeria. *Soc. Sci. Med. Mental Health*, 2, Article 100093.
- Pelham W, Gnagy E, Greenslade K, & Milich R (1992). Teachers’ rating of DSM-III-R symptoms for disruptive behavior disorders. *J. American Acad. Child and Adolescence Psych*, 31, 210–218.
- Putnick DL, & Bornstein MH (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev*, 41, 71–90. [PubMed: 27942093]
- Reise SP, & Waller NG (2009). Item response theory and clinical measurement. *Annu. Rev. Clin. Psychol*, 5, 27–48. [PubMed: 18976138]
- Reynolds CF, & Patel V (2017). Screening for depression: the global mental health context. *World Psychiatr*, 16(3), 316–317.
- Ryder AG, Yang J, Zhu X, Yao S, Yi J, Heine SJ, & Bagby RM (2008). The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? *J. Abnorm. Psychol*, 117(2), 300. [PubMed: 18489206]
- Simon GE, VonKorff M, Piccinelli M, Fullerton C, & Ormel J (1999). An international study of the relation between somatic symptoms and depression. *N. Engl. J. Med.*, 341(18), 1329–1335. [PubMed: 10536124]
- Simons G, & Fennig C (Eds.). (2018). *Ethnologue: Languages of the World, Twenty-First Edition.* Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Šimundi AM (2009). Measures of diagnostic accuracy: basic definitions. *EJIFCC*, 19(4), 203. [PubMed: 27683318]
- Super C, & Harkness S (1986). The developmental niche: a conceptualization at the interface of child and culture. *Int. J. Behav. Dev.*, 9(4), 545–569.
- Uneze A (2010). January 29). Nigeria: Number of Orphans, Vulnerable Children Hits 17.5 Million This Day. Retrieved <http://allafrica.com/stories/201001290366.html>. (Accessed 3 December 2018).

- Van Ommeren M, Sharma B, Thapa S, Makaju R, Prasain D, Bhattarai R, & de Jong J (1999). Preparing instruments for transcultural research: use of the translation monitoring form with Nepali-speaking Bhutanese refugees. *Transcult. Psychiatr*, 36(3), 285–301.
- Ventevogel P, Komproe I, Jordans M, Feo P, & De Jong J (2014). Validation of the Kirundi versions of brief self-rating scales for common mental disorders among children in Burundi. *BMC Psychiatr*, 14(1), 36.
- Watson LK, Kaiser BN, Giusto AM, Ayuku D, & Puffer ES (2019). Validating mental health assessment in Kenya using an innovative gold standard. *Int. J. Psychol e-pub ahead of print*.
- Weaver LJ, & Kaiser BN (2015). Developing and testing locally derived mental health scales: examples from North India and Haiti. *Field Methods*, 27(2), 115–130.
- Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, & Prince M (2009). The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. *J. Affect. Disord*, 113(1–2), 109–117. [PubMed: 18614241]

**Table 1**

Characteristics of participants (N = 330).

Characteristic	Hausa (n = 194, 58.8%)	Pidgin (n = 136, 41.2%)
	n (%)	n (%)
Age, mean (SD)	15.1 (1.7)	14.6 (1.7)
Gender (female)	40 (20.6)	44 (32.4)
Family size, mean (SD)	9.4 (5.1)	6.9 (3.0)
Living arrangement		
With both parents	132 (68.0)	108 (79.4)
Other	62 (32.0)	28 (20.6)
Education		
Primary or none	100 (51.8)	89 (65.4)
Secondary or more	93 (48.2)	47 (34.6)
Distressing event (PTSD)		
Death	116 (59.8)	73 (53.7)
Accident/Violence	35 (18.0)	20 (14.7)
Injury/Illness/Abuse	24 (12.4)	18 (13.2)
Disrupted household	6 (3.1)	8 (5.9)
Other	11 (5.7)	8 (5.9)
None	2 (1.0)	9 (6.6)

**Table 2**  
 Mean scores on Depression Self Rating Scale (DSRS) by language and clinical caseness (N = 327).

Item	English item	Keep	t-tests by language		t-tests by caseness	
			Hausa n = 194	Pidgin n = 136	Non-case n = 308	Case n = 21
DSRS-1r	Look forward		0.86	0.95	0.88	1.04
DSRS-2r	Sleep well		0.73	0.54	0.64	0.76
DSRS-3	Crying	Yes	0.55	0.49	0.49	0.95
DSRS-4r	Play		0.97	0.79	0.89	1.00
DSRS-5	Run away	Yes	0.33	0.37	0.33	0.62
DSRS-6	Tummy aches	Yes	0.38	0.54	0.43	0.62
DSRS-7r	Low energy	Yes	0.83	0.73	0.77	1.05
DSRS-8r	Enjoy food	Yes	0.61	0.46	0.54	0.81
DSRS-9r	Stuck up for self		0.74	0.70	0.72	0.76
DSRS-10	Isn't worth living	Yes	0.55	0.57	0.54	0.81
DSRS-11r	Good at things	Yes	0.79	0.57	0.67	1.05
DSRS-12r	Enjoy things	Yes	0.76	0.96	0.83	1.14
DSRS-13r	Like talking w/family	Yes	0.50	0.53	0.49	0.76
DSRS-14	Bad dreams	Yes	0.45	0.40	0.42	0.62
DSRS-15	Lonely	Yes	0.56	0.47	0.48	1.05
DSRS-16r	Easily cheered	Yes	0.96	0.75	0.86	1.14
DSRS-17	Sad	Yes	0.59	0.57	0.56	0.86
DSRS-18	Bored	Yes	0.69	0.72	0.68	0.95
DSRS-full	18 items		11.84	11.06	11.20	16.00
DSRS-cut	14 items		8.54	8.05	8.05	12.43
Cases			8%	4%	ns	ns

r = positive item recoded so that higher score reflects more negative response.

\* p < 0.20.

\*\* p < 0.05.

\*\*\* p < 0.01.

**Table 3**  
Mean scores on Child PTSD Symptom Scale (CPSS) by language and clinical caseness (N = 327).

Item	English item	Keep	t-tests by language			t-tests by caseness		
			Hausa n = 194	Pidgin n = 136	p	Non-case n = 289	Case n = 41	p
CPSS-1	Upsetting thoughts	Yes	2.1	2.3		2.2	2.5	**
CPSS-2	Bad dreams	Yes	1.6	1.5		1.5	1.8	*
CPSS-3	Act/feel as if happening again		1.7	1.7		1.7	1.8	
CPSS-4	Upset think about		2.6	2.4	**	2.5	2.7	
CPSS-5	Feelings in body		2.4	2.2	**	2.3	2.4	
CPSS-6	Try not to think about	Yes	2.3	2.3		2.2	2.5	*
CPSS-7	Avoid activities		2.1	2.2		2.1	2.4	
CPSS-8	Can't remember		2.3	2.0	**	2.2	2.3	
CPSS-9	Less interest		2.0	2.0		2.0	2.2	
CPSS-10r	Feeling close		2.8	2.6	*	2.8	2.5	
CPSS-11	No strong feelings	Yes	1.7	1.8		1.7	2.2	***
CPSS-12	Plans won't come true	Yes	1.7	1.6		1.6	2.0	***
CPSS-13	Trouble sleeping		1.7	1.8		1.7	1.8	
CPSS-14	Irritable	Yes	2.2	2.0	*	2.1	2.3	*
CPSS-15	Trouble concentrating	Yes	1.8	1.7		1.7	2.1	**
CPSS-16	Overly careful		2.6	2.5		2.6	2.6	
CPSS-17	Jumpy	Yes	2.0	2.2		2.0	2.6	***
CPSS-full	17 items		35.7	34.7	ns	34.8	38.6	***
CPSS-cut	8 items		15.4	15.4	ns			
Cases			15%	8%	**			

r = positive item recoded so that higher score reflects more negative response.

\* p < 0.20.

\*\* p < 0.05.

\*\*\* p < 0.01.

**Table 4**

Mean scores on Disruptive Behavior Disorder Rating Scale (DBDRS) – Oppositional Defiant Disorder subscale by language and clinical caseness (N = 328).

Item	English item	Keep	t-tests by language			t-tests by caseness (combined)			t-tests by caseness (Hausa)			t-tests by caseness (Pidgin)		
			Hausa n = 194	Pidgin n = 136	p	Non-case n = 271	Case n = 59	p	Non-case n = 159	Case n = 33	p	Non-case n = 111	Case n = 25	p
DBDRS-1	Loses temper	Yes	2.0	2.0	ns	2.0	2.3	**	2.0	2.2	*	1.9	2.4	**
DBDRS-2	Argues with adults	Yes	1.6	1.6	ns	1.5	1.9	***	1.6	1.8	*	1.5	2.0	**
DBDRS-3	Defies adults		2.2	2.9	***	2.5	2.4	ns	2.2	2.2	ns	3.0	2.5	**
DBDRS-4	Annoys people	Yes	1.5	1.5	ns	1.4	1.7	**	1.4	1.9	***	1.5	1.4	*
DBDRS-5	Blames others	Yes	1.4	1.6	ns	1.4	1.9	***	1.3	1.9	***	1.5	1.8	*
DBDRS-6	Easily annoyed	Yes	2.2	2.0	ns	2.0	2.4	***	2.1	2.5	**	2.0	2.4	*
DBDRS-7	Angry/resentful	Yes	2.0	2.1	ns	2.0	2.3	**	1.9	2.2	ns	2.0	2.7	**
DBDRS-8	Vindictive	Yes	1.6	1.3	***	1.5	1.7	**	1.6	1.8	ns	1.3	1.8	***
ODD-full	8 items		14.6	15.0	ns	14.3	16.7	***	14.1	16.5	***	14.6	17.0	***
ODD-cut	7 items		12.4	12.1	ns	11.8	14.3	***	10.3	12.5	***	10.3	12.7	***
ODD Cases			17%	18%	ns									

\* p < 0.20.

\*\* p < 0.05.

\*\*\* p < 0.01.



**Table 5**

Mean scores on Disruptive Behavior Disorder Rating Scale (DBDRS) – Conduct Disorder subscale by language and clinical caseness (N = 328).

Item	English item	Keep	t-tests by language			t-tests by caseness (combined)			t-tests by caseness (Hausa)			t-tests by caseness (Pidgin)		
			Hausa n = 194	Pidgin n = 136	p	Non-case n = 238	Case n = 92	p	Non-case n = 131	Case n = 61	p	Non-case n = 106	Case n = 30	p
DBDRS-9	Bullied, threatened	Yes	0.35	0.17	***	0.18	0.51	***	0.26	0.56	***	0.09	0.43	***
DBDRS-10	Initiated fights	Yes	0.39	0.32		0.29	0.57	***	0.27	0.66	***	0.30	0.40	***
DBDRS-11	Weapon	Yes	0.19	0.25		0.14	0.40	***	0.09	0.39	***	0.20	0.43	**
DBDRS-12	Physically cruel ppl	Yes	0.17	0.17		0.11	0.32	***	0.10	0.33	***	0.13	0.30	*
DBDRS-13	Physically cruel animals		0.15	0.26	**	0.16	0.28	**	0.09	0.28	***	0.24	0.33	***
DBDRS-14	Robbery w/ confrontation	Yes	0.10	0.10		0.05	0.25	***	0.05	0.23	***	0.05	0.30	***
DBDRS-15	Forced sex	Yes	0.07	0.01	***	0.01	0.13	***	0.02	0.18	***	0	0.03	
DBDRS-16	Fire setting		0.02	0.07		0.03	0.07		0.02	0.03		0.05	0.13	*
DBDRS-17	Destroyed property	Yes	0.15	0.10		0.08	0.26	***	0.06	0.34	***	0.10	0.10	
DBDRS-18	Broken into		0.05	0.04		0.03	0.07		0.03	0.08	*	0.04	0.03	
DBDRS-19	Cons others	Yes	0.47	0.56		0.43	0.72	***	0.37	0.70	***	0.49	0.80	***
DBDRS-20	Robbery w/o confrontation		0.25	0.36	**	0.21	0.51	***	0.13	0.52	***	0.32	0.50	*
DBDRS-21	Stays out	Yes	0.48	0.38		0.39	0.59	***	0.39	0.70	***	0.38	0.40	
DBDRS-22	Run away	Yes	0.21	0.12	**	0.11	0.34	***	0.11	0.43	***	0.10	0.17	
DBDRS-23	Truant	Yes	0.35	0.11	***	0.15	0.51	***	0.18	0.69	***	0.09	0.17	
CD-full	15 items		3.4	3.0	ns	2.4	5.5	***	2.2	6.1	***	2.6	4.5	***
CD-cut	11 items		2.9	2.3	**	1.9	4.6	***	1.9	5.2	***	1.9	3.5	***
CD Cases			32%	22%	**									

\* p < 0.20.

\*\* p < 0.05.

\*\*\* p < 0.01.

Psychometric properties of screening tools<sup>a</sup> compared to K-SADS<sup>b</sup> clinical caseness, N = 328.

**Table 6**

<b>6a. Depression Self Rating Scale</b>									
Cut-off score	Sens	Spec	PPV	NPV	LR+	LR-	DOR	YJ	
7	86	39	9	98	1.4	0.4	3.9	25	
8	86	45	10	98	1.5	0.3	4.8	30	
9	71	54	10	97	1.6	0.5	2.9	25	
<b>6b. Child PTSD Symptom Scale</b>									
Cut-off score	Sens	Spec	PPV	NPV	LR+	LR-	DOR	YJ	
12	95	26	15	97	1.3	0.2	7.0	21	
13	88	36	16	95	1.4	0.3	4.1	24	
14	80	43	17	94	1.4	0.5	3.1	24	
<b>6c. Disruptive Behavior Disorder Rating Scale – Oppositional Defiant Disorder subscale</b>									
Cut-off score	Sens	Spec	PPV	NPV	LR+	LR-	DOR	YJ	
10	97	24	22	97	1.3	0.1	9.2	21	
11	85	39	23	92	1.4	0.4	3.6	24	
12	73	52	25	90	1.5	0.5	2.9	25	
<b>6d. Disruptive Behavior Disorder Rating Scale – Conduct Disorder subscale</b>									
Cut-off score	Sens	Spec	PPV	NPV	LR+	LR-	DOR	YJ	
1	91	30	33	90	1.3	0.3	4.5	21	
2	83	51	40	88	1.7	0.3	5.0	34	
3	70	68	45	85	2.2	0.4	4.8	37	

<sup>a</sup> Rows highlighted in gray indicate the final cut-off score selected for each screening tool.

<sup>b</sup> Abbreviations: K-SADS: Kiddie-Schedule for Affective Disorders and Schizophrenia, PPV: positive predictive value, NPV: negative predictive value, LR+ : positive likelihood odds ratio, LR- : negative likelihood odds ratio, DOR: diagnostic odds ratio, YJ: Youden's index