Development process of an assessment tool for disruptive behavior problems in cross-cultural settings: the Disruptive Behavior International Scale – Nepal version (DBIS-N)

Matthew D. Burkey[1], **Lajina Ghimire**[2], **Ramesh P. Adhikari**[3], **Brandon A. Kohrt**[4], **Mark J. D. Jordans**[5], **Emily Haroz**[6], and **Lawrence Wissow**[7]

[1]Johns Hopkins University School of Public Health, Department of Health, Behavior, and Society, 624 N. Broadway, Baltimore, MD 21205, USA. Tel 250-392-4481

[2]Transcultural Psychosocial Organization (TPO) Nepal, PO Box 8974, Baluwatar, Kathmandu, Nepal. Tel +977-01-4431717

[3]Transcultural Psychosocial Organization (TPO) Nepal, PO Box 8974, Baluwatar, Kathmandu, Nepal. Tel +977-01-4431717

[4]Duke University, Duke Global Health Institute & Department of Psychiatry and Behavioral Sciences, 213 Trent Hall, 310 Trent Drive, Duke Box #90519, Durham, NC 27008, USA. Tel (919) 681-7516

[5]War Child, Helmholtzstraat 61-G, 1098 LE Amsterdam, Amsterdam, Netherlands. Tel +31-20-422-7777

[6]Johns Hopkins Bloomberg School of Public Health, Department of Mental Health, 624 N. Broadway, Baltimore, MD 21205, USA. Tel 410-955-3910

[7]Johns Hopkins Bloomberg School of Public Health, Department of Health, Behavior and Society, 624 N. Broadway, Baltimore, MD 21205, USA. Tel 410-614-1243

Abstract

Systematic processes are needed to develop valid measurement instruments for disruptive behavior disorders (DBDs) in cross-cultural settings. We employed a four-step process in Nepal to identify and select items for a culturally valid assessment instrument: 1) We extracted items from validated scales and local free-list interviews. 2) Parents, teachers, and peers (n=30) rated the perceived relevance and importance of behavior problems. 3) Highly rated items were piloted with children (n=60) in Nepal. 4) We evaluated internal consistency of the final scale. We identified 49 symptoms from 11 scales, and 39 behavior problems from free-list interviews (n=72). After dropping items for low ratings of relevance and severity and for poor item-test correlation, low frequency, and/or poor acceptability in pilot testing, 16 items remained for the Disruptive Behavior International Scale—Nepali version (DBIS-N). The final scale had good internal consistency

CORRESPONDENCE TO: Matthew Burkey, MD, PhD, 624 N. Broadway, Hampton House 703, Baltimore, MD 21205, USA. Tel: 250-392-4481. Fax: 410-955-8096. mburkey1@jhmi.edu.
lajina538@gmail.com, radhikari@tponepal.org.np, brandon.kohrt@duke.edu, mark.jordans@warchild.nl, eharoz1@jhu.edu, lwissow@jhsph.edu

(α=0.86). A 4-step systematic approach to scale development including local participation yielded an internally consistent scale that included culturally relevant behavior problems.

INTRODUCTION

Disruptive behavior disorders (DBDs) are among the most common child mental disorders and are important risk factors for academic failure, psychopathology, substance abuse, delinquency, and incarceration (Loeber, Burke, Lahey, Winters, & Zera, 2000). Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association, 2013) characterizes DBDs as patterns of behavior that "bring the individual into significant conflict with societal norms or authority figures". However, societal norms for child behavior vary between sociocultural settings (Harkness & Super, 1996; LeVine, 1988; Super et al., 2008; Whiting & Edwards, 1992), posing an important challenge for measuring child psychopathology, including DBDs (Canino & Alegría, 2008; Collins et al., 2011; Folmar & Palmes, 2009; B. A. Kohrt et al., 2011). Commonly used scales to measure DBDs in global mental health have largely been developed and validated in North America and Western Europe (Canino, Polanczyk, Bauermeister, Rohde, & Frick, 2010). Typically used translation/back-translation methods for scale adaptation carry the risk of reifying Western psychiatric categories in settings where the constructs are devoid of meaning (i.e., "category fallacy" (Kleinman, 1987))(B. A. Kohrt et al., 2011). While quantitative evaluations may fail to identify a category fallacy, previous studies in non-Western settings have demonstrated variability in average externalizing symptom scores between cultural settings (Crijnen, Achenbach, & Verhulst, 1997) and that diagnoses of Oppositional Defiant Disorder and Conduct Disorder by structured clinical interviews failed to converge with local behavior problem concepts (Ng, Kanyanganzi, Munyanah, Mushashi, & Betancourt, 2014). In Nepal, a large study identified a distinct factor structure for the Conners Scale (Pendergast, 2013) and another study documented an unsuccessful attempt to translate ADHD symptoms into Nepali due to lack of coherence of symptoms in the local context (Folmar & Palmes, 2009).

In order to maximize the content validity of DBD measurement tools and reduce the risk of category fallacy, systematic procedures are needed to account for cross-cultural variation in societal norms for child behavior. This paper evaluates a procedure that utilizes local participants' insider ("emic") knowledge of child behavioral expectations to identify and select items for measuring DBDs in cross-cultural settings.

Advances in development of scales with locally derived content have come from the increasing use of free-listing interviews with beneficiary cultural groups to generate and select salient items (Betancourt et al., 2009; Bolton & Tang, 2002; Ng et al., 2014). Free-listing is a qualitative interview technique used by cultural anthropologists and others to describe semantic networks within cultural domains (Borgatti, 1999) and provides a useful measures of "salience" and "prototypicality" (Thompson & Juan, 2006). Improvements on

these methods need to address two limitations of free-listing relevant to scale development: 1) item pools derived exclusively from free-listing are often small and lack completeness, and 2) other techniques are better suited to assessing severity-related relevance.

Standard scale development guidelines suggest developing a large pool of candidate items – 3- to 4-times as large as the anticipated length of the final scale – that represent the construct of interest as completely as possible (DeVellis, 2011). A large initial item pool covers the breadth of the target construct and facilitates dropping less relevant or poorly performing items at later stages. However, free-list interviews often lead to limited sets of responses (Brewer, 2002). Previous studies using free-listing to generate items for behavior problem-related scales in LMIC have started with pools of 13 or 21 items (Betancourt et al., 2009; Ng et al., 2014). Moreover, the content of free-list interviews may be biased by the prompt provided, which may not capture all relevant elements. Given these limitations, advances are needed that expand the size and completeness of the initial item pools.

In addition, after a large item pool has been developed, DeVellis (2011) recommends that content experts review the items to aid in the selection of those that are most relevant to the target construct. While frequency of mention in free-lists is often taken as a proxy for relevance, DeVellis (pg. 86) recommends using content experts' direct ratings of items' importance and relevance. Somewhat differently than for other mental health constructs, the relevance of behavior problem items can be conceptualized as a function of the degree to which behaviors violate societal norms (American Psychiatric Association, 2013). This ('emic') knowledge of perceived severity is possessed by those who--by their evaluations and responses in everyday life—define and reinforce local behavioral norms. Therefore, comparative rating methods (such as Likert scales) with local stakeholders may be well suited for evaluating the relevance of behavior problems vis-à-vis local behavioral norms.

We propose a novel procedure for scale development for DBDs in cross-cultural settings that addresses the shortcomings of current free-listing-based methods by: 1) generating a large initial item pool integrating items from local free-list interviews and existing validated scales; and, 2) narrowing items for pilot testing using direct ratings of perceived severity by individuals with emic knowledge of local behavioral norms. This procedure has the benefit of capturing items that may be missed in free-listing but that local experts rate as important when introduced from existing scales.

In this paper, we describe the 4-step process used to develop a scale rating disruptive behavior problems among children and youth in Nepal. To demonstrate that the method addresses the difficulties described above, we hypothesized that it would result in a larger initial item pool than previous scale development efforts for behavior problems that have used free-listing alone. Second, we hypothesized that some items derived from existing scales but not mentioned in local free-list interviews would be rated highly (i.e. in the top quartile) by local stakeholders on criteria of importance and relevance.

METHODS

Study Context

We developed the Disruptive Behavior International Scale—Nepal version (DBIS-N) within a broader study of child behavior problems in Nepal, a low-income country in South Asia. Nepal has high rates of extreme poverty, child malnutrition, and migration and recent high exposure to conflict during the People's War in Nepal (1996-2006) (UNICEF, 2006). As part of an ongoing project aiming to establish mental health care in Nepal (Jordans, Luitel, Pokharel, & Patel, 2015), our team has conducted formative research in order to understand stakeholders' concerns related to child behavior problems and effectively target an intervention toward locally meaningful and acceptable goals (Burkey et al., 2015).

Purpose of the Instrument

The primary purpose of the instrument developed in this paper was to identify children with behavior-related problems who might benefit from a treatment intervention. The construct we sought to measure was behavior-related problems in children that were broadly related to disruptive, aggressive, and/or antisocial behaviors. In order to maintain relevance to existing empirical literature, our guiding construct was based largely on the broad category of Disruptive Behavior Disorders in DSM-5 (American Psychiatric Association, 2013). We also remained open to local concerns and priorities in order to reduce the possibility of reifying a disorder construct devoid of local coherence (i.e. "category fallacy" (Kleinman, 1987)).

Ethics Statement

This study was approved by the Johns Hopkins Bloomberg School of Public Health IRB and the Nepal Health Research Council. All participants provided informed consent (and children provided assent) and were compensated for their time.

Step 1: Item generation

To generate a pool of behavior-related problems from which to develop a locally adapted tool, we used both free-list interviews and a review of existing tools. We began by conducting free-list interviews with teachers and parents in the local community in Nepal (total N=72) (Adhikari et al., 2015). Each participant was asked: "Please tell us about the problems children between 8-15 years are facing in your community." We coded behavior-related problems and tabulated the frequency of each. Interviews and coding were conducted in Nepali and then translated into English. Problems were included as items in this study if they were mentioned by at least 3 respondents. We excluded problems related to socioeconomic conditions.

Next, we sought to add items from existing instruments that measure DBD-related constructs. We identified instruments by searching MEDLINE and PsycINFO and by hand-searching references and web resources. We included instruments that evaluated DBD-related constructs (including ODD, CD, aggression, or closely related disruptive behavior problems) with at least one positive measure of concurrent or criterion validity reported in a peer-reviewed published report that included at least 100 subjects. Instruments were excluded if they evaluated only adults (over 18 years). We then coded and extracted items

using NVivo (QSR International, 2012), grouping items by conceptual similarity and tabulating the frequency of each symptom.

Step 2: Item relevance

We then translated each item into Nepali and assessed the comprehensibility, importance, and relevance of each item to potential respondents and key stakeholders. We assessed comprehensibility in two focus groups of parents and teachers using probing questions to identify and resolve potential barriers to understandability. A bilingual Nepali-English speaker blinded to the instrument then back-translated the modified items into English to check for conceptual equivalence.

We then assessed the importance (i.e. perceived severity) and relevance of each item using a structured survey with 10 children (ages 8-15, i.e. "peer perspective"), 10 teachers, and 10 parents (50% female in each category). The framework for assessing item importance was based on our previous ethnographic research in Nepal indicating that a widely shared and highly valued desire among parents is to ensure a "bright future" (Nepali: *ujjwala bhavishya*) and avoid a "dark future" (*amdhyaro bhayishya*) for their children (Burkey et al., 2015). Each respondent rated importance on a 1-to-4 scale ('4' represented behaviors most likely to lead to a dark future.) Each respondent also rated the relevance of each item to the local terminology related to bad behavior (*badmaash*) ('4' indicated behaviors most indicative of *badmaash*.)

We then selected the items for piloting in the next step based on criteria of comprehensibility, importance, and relevance. We also included a small number of items with lower importance/relevance ratings for piloting if they were included in a majority of validated scales in order to include items with potential global significance.

Step 3: Item utility

To assess the performance of individual items in situations resembling actual usage, we then pilot tested the narrowed set of items in a "development sample" of children in the local community (DeVellis, 2011). Respondents in the development sample were parents of children aged 5-15, selected using a convenience sample of households in the target community. Response options included: 0—"Never/rarely", 1—"Occasionally", and 2—"Often". During pilot testing interviews, the research assistants also took notes concerning parents' difficulty understanding questions and barriers to acceptability of asking the questions.

Following pilot testing, we dropped items for the final scale based on the following criteria: 1) lack of acceptability of asking the item (based on solicited feedback from parents); 2) low item-test correlation (Pearson's correlation coefficient <0.20); and/or 3) extremes of frequency (i.e. item was rated as highest or lowest response choice in >80% of those sampled.) We also selected a subset of items that would only be asked for older children (10-15 year olds), given considerations about local epidemiologic patterns and acceptability of asking questions about serious offenses of younger children.

Step 4: Psychometric properties and evaluation of procedures

The goal of step 4 was to conduct an initial evaluation of the psychometric properties of the scale and to evaluate the utility of adding items from existing scales to the initial item pool and of using stakeholder ratings as a method to select items for the scale. We assessed the internal consistency of the resulting scale using Cronbach's alpha to analyze data from the development sample.

We evaluated the utility of including items from both free-list interviews and existing scales by tabulating the number of unique items generated from each method, comparing the mean importance and relevance ratings of items from each source using t-tests, and evaluating the source of items rated in the top quartile for relevance and importance. Statistical analyses were performed in Stata 12 (Stata Corporation, 1985-2013).

RESULTS

Step 1: Item Generation

Free-list interviews (n=72) with local parents and teachers revealed 39 unique behavior problems of concern to at least 3 respondents. Free-list participants were particularly concerned about local patterns of "bad habits" (e.g. gambling), maintaining hygiene (i.e. washing), sexual mores (e.g. proper dress, watching pornography, premature interest in dating), and leisure activities (e.g. watching TV or using cell phones too much).

We identified 11 published instruments that met inclusion criteria for review (Supplemental Table 1). Of these, ten were developed in the United States or Western Europe and one in East Africa (Ng et al., 2014). The included scales varied in length and measured a variety of DBD-related constructs. We identified 218 items from the scales that related to disruptive behavior problems, representing 49 unique symptoms. The final pool consisted of 62 unique symptoms comprised of 13 items (21%) from free-listing, 23 items (37%) from existing tools, and 26 items (42%) from both sources (i.e. overlapping). Figure 2 illustrates the identification and selection of items throughout the study's three phases.

Step 2: Item relevance

Participants in focus group discussions (n=10) identified problems with items' comprehensibility and suggested improvements. Problems with comprehensibility largely related to difficulty understanding the terms and phrases used to describe behaviors. For example, "watching pornography films" was not understood by several elderly respondents. A local term, "blue films" (spoken in English) (also used in other parts of South Asia) was better understood by local participants, but remained unfamiliar to many. Participants noted problems with the relevance of items like "beating animals", which was associated with common animal herding practices and not viewed as a problematic behavior.

Local stakeholders (child peers, parents, and teachers) rated the 62 candidate items for importance (i.e. association with "dark future") and relevance (i.e. to local behavior problem term of *badmaash*) (see Table 1). Importance and relevance were strongly correlated (Spearman's rho = 0.87, p=<0.0001). Mean importance rating was 2.88 and mean relevance

rating was 2.84. The highest-rated items (combined score) were: using a dangerous weapon on others, smoking marijuana, stealing from non-family members, and *drinking alcohol*. The lowest-rated items were: roaming around or wandering, watching TV too much, acting "mischievous" (Nepali: *chakchake*)*, using mobile phone too much,* and *not sharing*.

In total, 32 items were selected for pilot testing on the basis of comprehensibility, acceptability, and at least one indicator of importance: importance/relevance ratings (n=27), inclusion in a majority of reviewed scales (n=4), and prominence in qualitative interviews in the local community (n=1) (Burkey et al., 2015).

Step 3: Item utility

These administered the 32-item version of the tool to the parents of 60 children in the local community. The children in this development sample had a mean age of 10.2 (SD: 3.2, range: 5-15) and 60% were female. Results in the development sample and comments from parents highlighted additional problems with some items related to low frequency, poor item-test correlation, poor comprehensibility in test settings, and poor acceptability. For example, *using a dangerous weapon* and *deliberately setting fires to cause damage* were rated as "Never/Rarely" by 97% of respondents and were dropped. Multiple attempts were made to identify an equivalent translation for *temper tantrums* but none was widely understood. Multiple parents stated they felt uncomfortable when asked about substance use or sexual behavior in their younger children, especially young girls. After dropping problematic items, sixteen questions remained for the final instrument.

Based on feedback from parent respondents who were concerned about the deficit-focused questions in the pilot phase, we added 4 items to assess pro-social child behaviors derived from recent qualitative interviews with local stakeholders (Burkey et al, 2016). We also modified the response choices to include 4 options ("Never", "Sometimes", "Often", and "Very Often") in order to enhance precision and increase variability in responses. The final version of the instrument included 16 problem items, 4 pro-social items, and a 4-item supplement for older children and adolescents (Figure 3.) Items in the adolescent supplement address widespread concerns about substance use/abuse and running away which were statistically infrequent and culturally inappropriate to ask younger children.

Phase 4: Psychometric properties and evaluation of procedures

Cronbach's alpha based on parents in the development sample rating their children on 16 problem items was 0.86 (N=60).

Compared with items identified from free-listing alone (n=13), items identified from existing scales alone (n=23) were rated slightly higher for importance (mean: 2.88 vs. 2.62, t(34)=2.23, p=0.03) but not relevance (mean: 2.90 vs 2.77, t(34)=1.02, p=0.31). Among items rated in the top quartile for importance, 5 came from existing scales alone, 3 were from free-listing alone, and 5 were from both sources. In the top quartile for relevance, 5 items came from existing scales alone, 1 from free-listing alone, and 7 from both sources.

DISCUSSION

While many instruments have been validated for the measurement of disruptive behavior problems, only a few have been developed outside of North America or Europe. This paper describes the application of a systematic procedure to incorporate local stakeholder participation for generating and selecting items for the <u>D</u>isruptive <u>B</u>ehavior <u>I</u>nternational <u>S</u>cale—<u>N</u>epal version (DBIS-N). Using local free-list interviews and a review of existing scales, we identified 68 unique behavior problem items. Based on these local stakeholders' ratings of importance and relevance, the initial pool was narrowed to 32 items for pilot testing. Results from the development sample informed the selection of 16 behavior problem items that were found to be moderately prevalent, acceptable to ask, and internally reliable. As hypothesized, combining free-lists and existing tools to generate candidate items resulted in a large number of items in the initial pool, and items from both local and external sources received high ratings of importance and relevance. Our findings support drawing upon internally and externally derived item pools and the use of local participation as an efficient, and potentially widely applicable, component of scale development to address cross-cultural variation in DBDs.

The process we used demonstrates an adaptation of DeVellis' (2011) framework for scale development that may be useful in other global mental health settings. Our study highlights the utility of using local informants as "experts" in disorder constructs that closely relate to local behavioral expectations—in this case, child behavior problems. In our study, parents and teachers from the local community provided feedback on acceptability as well as the relevance and importance of candidate items assessing child behavior problems. Their feedback helped narrow a large initial item pool into a smaller set of items that could more feasibly and efficiently be assessed in pilot testing in a development sample. The high internal consistency of the final scale suggests that stakeholder participation helped to select items that measure a cohesive underlying construct.

Our evaluation of candidate items found a large degree of overlap between locally identified problem behaviors and items from externally derived scales. However, there was also a subset of symptoms we identified in the local context that were only shared by the single other study we found of an instrument developed in another LMIC (i.e. Rwanda) (Ng et al., 2014). Both our study and the study in Rwanda identified overlapping concerns that local residents identified as behavior problems but that are not commonly included in existing instruments. These include: *roaming around*, *speaking rudely*, *sexually deviant behavior*, *being impolite*, *taking drugs/alcohol*, *failing to maintain hygiene*, doing other "bad behaviors" not specified in DSM (e.g. gambling), and *being ungrateful*. While several of these items were dropped from the DBIS due to concerns about acceptability or lower importance ratings, these findings suggest that there may be sets of concerns that many parents in LMIC settings identify as behavior-related problems that are not represented by "Western" concepts of disruptive behavior disorders (as in the DSM (American Psychiatric Association, 2013)). These findings support the need for "ground up" approaches to scale development for child behavior problems in novel sociocultural contexts.

Given the context-specificity of child behavioral norms, we expected to find a smaller degree of overlap in symptoms between the locally-derived symptoms and the items from international scales. There may be a number of explanations for this observation. Some child behavioral patterns may be universally concerning (or nearly so) to peers, parents, and/or teachers. This may be especially true among teachers, given the international influences prominent in teachers' education (e.g. textbooks, urban-based education), and in school classrooms given a similar structure and demands across settings (Pope Edwards, Gandini, & Giovaninni, 1996).

An alternative explanation is that the overlap of symptoms represents a "category fallacy" (Kleinman, 1987). Kleinman noted that investigators looking for specific mental health syndromes in new sociocultural contexts may inadvertently 'reify' the syndromes they are looking for, but that these syndromes either lack coherence or have different meanings (Kleinman, 1987). Given this common methodical and conceptual failure of some cultural psychiatry and global mental health research, we did not attempt to identify a narrowly-defined syndrome (e.g. conduct disorder) and we do not treat *badmaash* as a syndrome. Instead, we used a "ground up" (inductive) approach that prioritized local concerns for child behavior (broadly defined) as a filter to select items derived from both local and international sources. In a separate paper (Burkey et al., 2015) we describe the qualitative research that identified a problem area of *badmaash* child behavior as well as locally meaningful frameworks used in this paper to assess importance and relevance.

Limitations

Our results concerning the psychometric properties of the scale and its items are from a "development phase" pilot that was exploratory in nature, and relied on a small convenience sample of children in a single community in Nepal. Conclusions about other reliability properties of the DBIS-N, construct validity of the DBIS-N, the relationship between pro-social and problem items, and the utility of the supplemental questions for older children await the results of an ongoing study in a larger validation sample in Nepal.

Additional questions remain about the transferability of stakeholder ratings to other sociocultural settings and disorder constructs. We note that incorporating laypersons' feedback on the importance of items may not be as useful when developing scales targeting a construct that is thought to depend less upon culture-specific behavioral norms (such as schizophrenia) or have more "universal" characteristics. However, even in these contexts, obtaining feedback from the population targeted by the instrument will help develop items phrased in ways that are comprehensible, acceptable, and relevant to local circumstances (Van Ommeren et al., 1999).

Applications

Applications of the DBIS include local epidemiological assessments, screening for interventions, and evaluating intervention outcomes. In addition, this systematically and locally developed tool may aid efforts in global mental health and neuroscience (Stein et al., 2015) to identify cross-cultural biological markers and mechanisms related to DBDs. Prior studies with boys in Nepal and other regions of Asia have demonstrated associations of

disruptive behaviors with hypocortisolism in naturalistic assessments (Hruschka, Kohrt, & Worthman, 2005; Brandon A Kohrt et al., 2015).

In addition, we anticipate that the process and results of this study will serve as a template for developing similar locally adapted instruments for DBDs in other contexts. Growing attention to the mental health needs of ethnic Nepali Bhutanese refugee youth in the United States and other high-income countries is an area of potential application. Recent research utilizing key informant interviews in addition to free lists identified similar concerns associated with youth characterized as *badmaash* (Betancourt et al., 2015). Descriptors identified in the Bhutanese youth study strongly overlapped with items generated in southern Nepal. In order to facilitate transferability of this process to other settings, we plan to make an extensible version of our item database (Step 1), structured data collection tools (Step 2), and data analysis coding and decision aids (Step 3) available online (through the Mental Health Innovation Network website (http://mhinnovation.net/)).

CONCLUSIONS

Valid instruments that assess parents' and teachers' concerns for child behavior problems are needed to identify children who would benefit from targeted treatment interventions. Instruments in common use were developed in high-income, Western settings; current adaptation procedures are limited by the lack of input from key stakeholders in child development and may fail to address important societal norms for child behavior. To address this gap, we developed the DBIS-N using procedures that incorporate local participation for item generation and selection. Our process requires further assessment of construct validity in Nepal and replication in other sociocultural settings to better characterize its transferability. Through the systematic development of tools that account for local concerns, we will better be able target interventions to the children that need them, measure interventions' effectiveness, and meet the needs of a culturally diverse population of children worldwide.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

Adhikari, Ramesh P.; Upadhaya, Nawaraj; Gurung, Dristy; Luitel, Nagendra P.; Burkey, Matthew D.; Kohrt, Brandon A.; Jordans, Mark JD. Perceived behavioral problems of school aged children in

rural Nepal: a qualitative study. Child and Adolescent Psychiatry and Mental Health. 2015; 9(25) doi 10.1186/s13034-015-0061-8.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders, DSM-5. Arlington, VA: 2013.

Betancourt, Theresa S.; Bass, Judith; Borisova, Ivelina; Neugebauer, Richard; Speelman, Liesbeth; Onyango, Grace; Bolton, Paul. Assessing local instrument reliability and validity: a field-based example from northern Uganda. Social Psychiatry and Psychiatric Epidemiology. 2009; 44(8):685–692. [PubMed: 19165403]

Bolton, Paul; Tang, Alice M. An alternative approach to cross-cultural function assessment. Social Psychiatry and Psychiatric Epidemiology. 2002; 37(11):537–543. [PubMed: 12395144]

Borgatti, Stephen P. Elicitation techniques for cultural domain analysis. In: Schensul, JJ.; LeCompte, MD., editors. Enhanced ethnographic methods: audiovisual techniques, focused group interviews, and elicitation techniques. Vol. 3. Altimira Press; Walnut Creek, CA: 1999. p. 1-26.

Brewer, Devon D. Supplementary interviewing techniques to maximize output in free listing tasks. Field methods. 2002; 14(1):108–118.

Burkey, Matthew D.; Ghimire, L.; Adhikari, R.; Luitel, Nagendra P.; Jordans, Mark J.; Wissow, Lawrence S.; Kohrt, Brandon A. Roles of Cultural Context in Definitions of and Responses to Conduct Problems in Children: A Case Study from Nepal. Paper presented at the Society for the Study of Psychiatry and Culture, Providence, RI. 2015

Canino, Glorisa; Alegría, Margarita. Psychiatric diagnosis–is it universal or relative to culture? Journal of Child Psychology and Psychiatry. 2008; 49(3):237–250. [PubMed: 18333929]

Canino, Glorisa; Polanczyk, Guilherme; Bauermeister, Jose J.; Rohde, Luis A.; Frick, Paul J. Does the prevalence of CD and ODD vary across cultures? Social Psychiatry and Psychiatric Epidemiology. 2010; 45(7):695–704. [PubMed: 20532864]

Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, Stein DJ. Grand challenges in global mental health. Nature. 2011; 475(7354):27–30. doi: 10.1038/475027a. [PubMed: 21734685]

Crijnen AA, Achenbach TM, Verhulst FC. Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36(9):1269–1277. doi: 10.1097/00004583-199709000-00020. [PubMed: 9291729]

DeVellis, Robert F. Scale development: Theory and applications. Vol. 26. Sage Publications; Los Angeles: 2011.

Folmar, Steven; Palmes, Guy. Cross-cultural psychiatry in the field: collaborating with anthropology. Journal of the American Academy of Child andAdolescent Psychiatry. 2009; 48(9):873.

Harkness, Sara; Super, Charles M. Parents cultural belief systems: Their origins, expressions, and consequences. Guilford Press; New York: 1996.

Hruschka DJ, Kohrt BA, Worthman CM. Estimating between- and within-individual variation in cortisol levels using multilevel models. Psychoneuroendocrinology. 2005; 30(7):698–714. doi: 10.1016/j.psyneuen.2005.03.002. [PubMed: 15854786]

Jordans MJ, Luitel NP, Pokharel P, Patel V. Development and pilot testing of a mental health care plan in Nepal. British Journal of Psychiatry. 2015 in press.

Kleinman, Arthur. Anthropology and psychiatry. The role of culture in cross-cultural research on illness. The British Journal of Psychiatry. 1987; 151(4):447–454. [PubMed: 3447661]

Kohrt BA, Jordans MJD, Tol WA, Luitel NP, Maharjan SM, Upadhaya N. 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 Psychiatry. 2011; 11(1):1–17. [PubMed: 21194496]

Kohrt, Brandon A.; Hruschka, Daniel J.; Kohrt, Holbrook E.; Carrion, Victor G.; Waldman, Irwin D.; Worthman, Carol M. Child abuse, disruptive behavior disorders, depression, and salivary cortisol levels among institutionalized and community-residing boys in Mongolia. Asia-Pacific Psychiatry. 2015; 7(1):7–19. [PubMed: 24890783]

LeVine, Robert A. Human parental care: Universal goals, cultural strategies, individual behavior. New Directions for Child and Adolescent Development. 1988; 1988(40):3–12.

Loeber, Rolf; Burke, Jeffrey D.; Lahey, Benjamin B.; Winters, Alaina; Zera, Marcie. Oppositional defiant and conduct disorder: a review of the past 10 years, part I. Journal of the American Academy of Child and Adolescent Psychiatry. 2000; 39(12):1468–1484. [PubMed: 11128323]

Ng, Lauren C.; Kanyanganzi, Frederick; Munyanah, Morris; Mushashi, Christine; Betancourt, Theresa S. Developing and Validating the Youth Conduct Problems Scale-Rwanda: A Mixed Methods Approach. PloS One. 2014; 9(6):e100549. [PubMed: 24949628]

Pendergast, Laura L. Factor structure of scores from the Conners' Parent and Teacher Rating Scales-Revised among children in Nepal. The Pennsylvania State University; 2013.

Pope Edwards, Carolyn; Gandini, Lella; Giovaninni, Donatella. The contrasting developmental timetables of parents and pre-school teachers in two cultural communities. In: Harkness, S.; Super, CM., editors. Parents cultural belief systems: Their origins, expressions, and consequences. Guilford Press; New York: 1996. p. 270-288.

QSR International. NVivo (Version 10). QSR International Pty Ltd; 2012.

Stata Corporation. Stata Statistical Software Release 12.0: Programming. Stata Corporation; College Station, Texas: 1985-2013.

Stein, Dan J.; He, Yanling; Phillips, Anthony; Sahakian, Barbara J.; Williams, John; Patel, Vikram. Global mental health and neuroscience: potential synergies. The Lancet Psychiatry. 2015; 2(2): 178–185. [PubMed: 26359754]

Super, Charles M.; Axia, Giovanna; Harkness, Sara; Welles-Nystrom, Barbara; Zylicz, Piotr Olaf; Parmar, Parminder; Kolar, Violet. Culture, temperament, and the "difficult child": A study in seven Western cultures. International Journal of Developmental Science. 2008; 2(1):136–157.

Thompson, Eric C.; Juan, Zhang. Comparative cultural salience: measures using free-list data. Field methods. 2006; 18(4):398–412.

UNICEF. Situation of children and women in Nepal. UNICEF; 2006.

Van Ommeren, Mark; Sharma, Bhogendra; Thapa, Suraj; Makaju, Ramesh; Prasain, Dinesh; Bhattarai, Rabindra; de Jong, Joop. Preparing instruments for transcultural research: use of the translation monitoring form with Nepali-speaking Bhutanese refugees. Transcultural Psychiatry. 1999; 36(3): 285–301.

Whiting, Beatrice B.; Edwards, Carolyn Pope. Children of different worlds: The formation of social behavior. Harvard University Press; 1992.
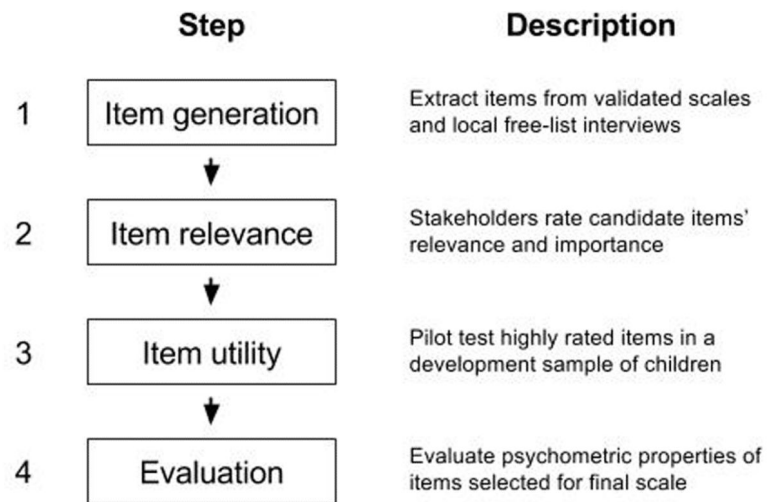
Figure 1.
Study flow diagram illustrating the study phases for generation, selection, and evaluation of items for the Disruptive Behavior International Scale—Nepal version (DBIS-N)
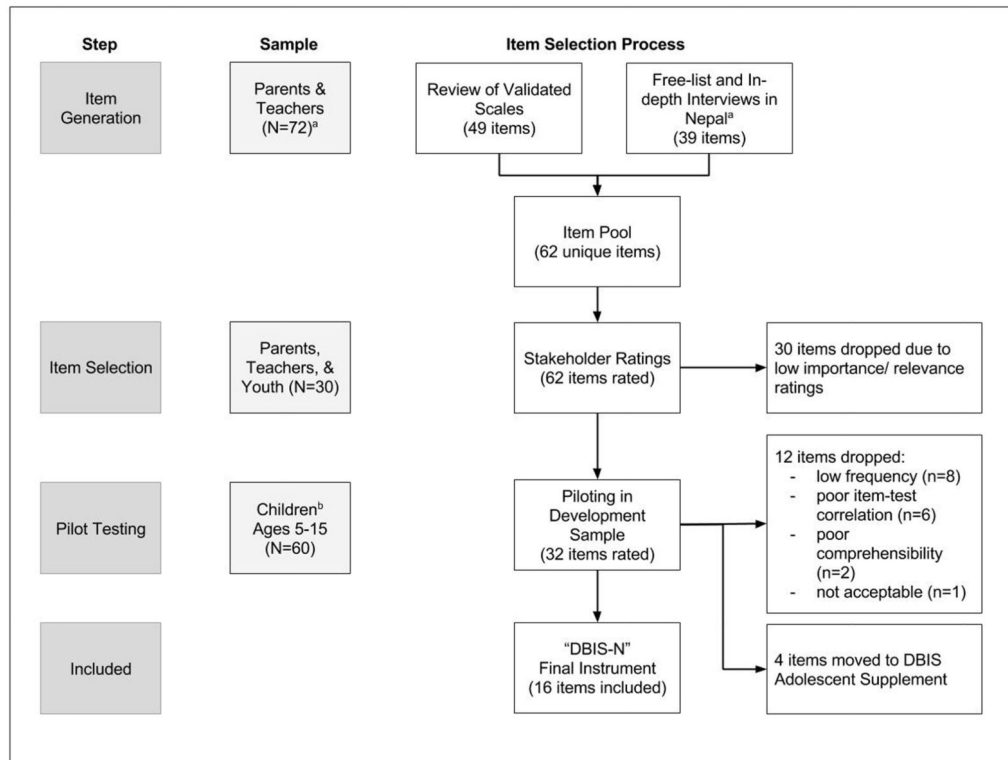
a Free-list interview results reported in Adhikari et al, 2015; In-depth interview results reported in Burkey et al, 2016
b Children's behavior problems were rated by a parent

Figure 2.
Study flow diagram illustrating the sources of items, and reasons items were dropped throughout the analysis and pilot testing phases, resulting in the Disruptive Behavior International Scale—Nepal version (DBIS-N). (Note: Some items were dropped due to more than one reason.)

Disruptive Behavior International Scale—Nepali Version

Instructions: Read each question to the child's primary caregiver. After reading the question, ask whether the child has done this "rarely, sometimes, often, or very often" over the past 6 months. Circle the best response (only one per question. Questions 21-24 are for children age 10-15 only.

Question	Response (Circle one)			
Before each question read: *"In the past 6 months…"*	**Never:** child never does this **Sometimes:** only once or twice per month **Often:** at least once per week **Very Often:** almost every day			
Pro-social Subscale				
1. How often did your child get along well with other children?	Never	Sometimes	Often	Very Often
2. How often did your child treat elders respectfully?	Never	Sometimes	Often	Very Often
3. How often did your child complete his/her schoolwork willingly?	Never	Sometimes	Often	Very Often
4. How often did your child help with household work?	Never	Sometimes	Often	Very Often
Problem Subscale				
5. How often was your child boldly disobedient?	Never	Sometimes	Often	Very Often
6. How often did your child get angry even on small things?	Never	Sometimes	Often	Very Often
7. How often did your child skip school?	Never	Sometimes	Often	Very Often
8. How often did your child curse or use foul words?	Never	Sometimes	Often	Very Often
9. How often did your child lie?	Never	Sometimes	Often	Very Often
10. How often did your child fail to follow instructions from elders?	Never	Sometimes	Often	Very Often
11. How often did your child fight with other children?	Never	Sometimes	Often	Very Often
12. How often did your child seek revenge?	Never	Sometimes	Often	Very Often
13. How often did your child spend time with children who do bad things? ("walks in bad circle")	Never	Sometimes	Often	Very Often
14. How often did your child do things to deliberately annoy others?	Never	Sometimes	Often	Very Often
15. How often did your child argue with elders?	Never	Sometimes	Often	Very Often
16. How often did your child damage or destroy others' property on purpose?	Never	Sometimes	Often	Very Often
17. How often did your child blame others for his/her own mistakes?	Never	Sometimes	Often	Very Often
18. How often did your child talk back to adults?	Never	Sometimes	Often	Very Often
19. How often did your child fail to pay attention to hygiene and cleanliness?	Never	Sometimes	Often	Very Often
20. How often did your child take valuable items from people outside the family?	Never	Sometimes	Often	Very Often
Questions for Older Children (ages 10-15)				
21. How often did your child run away from home overnight?	Never	Sometimes	Often	Very Often
22. How often did your child smoke marijuana?	Never	Sometimes	Often	Very Often
23. How often did your child use cigarettes or tobacco?	Never	Sometimes	Often	Very Often
24. How often did your child drink alcohol?	Never	Sometimes	Often	Very Often

Scoring: To score the subscales, add the scores for each item within the subscale, where "Never"=0, "Sometimes"=1, "Often"=2, and "Very often"=3

Figure 3.
The final version of the Disruptive Behavior International Scale—Nepal version (DBIS)

Table 1

Ratings of Perceived Relevance and Importance for 32 Behavior Problem Items Selected for Pilot Testing

Item		Relevance[1] Mean (SD)	Importance[2] Mean (SD)
1.	Has used a dangerous weapon on others	3.83 (0.38)	3.71 (0.66)
2.	Smokes marijuana	3.72 (0.45)	3.66 (0.61)
3.	Takes things (steals) from people outside the family	3.59 (0.63)	3.72 (0.59)
4.	Drinks alcohol	3.76 (0.44)	3.45 (0.69)
5.	Deliberately sets fires to cause damage	3.48 (0.69)	3.57 (0.74)
6.	Uses cigarettes or tobacco	3.55 (0.57)	3.31 (0.66)
7.	Carries a weapon	3.52 (0.63)	3.25 (0.80)
8.	Involved in physical relationship or watches porn movies	3.43 (0.74)	3.25 (1.04)
9.	Does dangerous things often	3.31 (0.66)	3.32 (0.77)
10.	Fights often	3.28 (0.70)	3.21 (0.79)
11.	Blames others for own mistakes	3.24 (0.83)	3.18 (0.86)
12.	Lies often	3.17 (0.71)	3.14 (0.74)
13.	Seeks revenge	3.17 (0.85)	3.11 (0.96)
14.	Gambles	3.21 (0.86)	3.07 (0.94)
15.	Argues with elders	3.07 (0.92)	3.14 (0.97)
16.	Wears improper or indecent clothing	3.24 (0.83)	2.86 (0.85)
17.	Takes things (steals) from family members without asking	3.00 (1.00)	3.10 (0.98)
18.	Boldly disobedient	3.07 (0.84)	3.03 (0.78)
19.	Damages or destroys others' property on purpose	2.97 (0.91)	3.11 (0.88)
20.	Spends time with children who do bad things (Walks in bad circle)	3.00 (0.76)	3.04 (0.96)
21.	Talks back to adults	3.00 (0.89)	3.04 (0.92)
22.	Runs away from home	3.07 (0.84)	2.97 (0.94)
23.	Does not follow rules (family rules)	3.17 (0.97)	2.86 (0.79)
24.	Curses or uses foul words	2.86 (0.95)	3.10 (0.82)
25.	Skips school	2.79 (1.05)	3.11 (0.83)
26.	Seeks attention from others too often	3.03 (0.78)	2.86 (0.80)
27.	Threatens others	2.86 (0.79)	2.96 (0.88)
28.	Doesn't pay attention to hygiene and cleanliness	2.79 (0.90)	2.61 (0.99)
29.	Does things to deliberately annoy others	2.79 (0.86)	2.54 (0.96)
30.	Harasses (teases or bullies) other children	2.59 (0.68)	2.61 (0.96)
31.	Has frequent temper tantrums (or anger outbursts)	2.62 (1.05)	2.43 (1.03)
32.	Gets angry even on small things	2.38 (1.01)	2.57 (0.74)

[1]**Relevance** was rated on a 1 to 4 scale: '1' Item not associated with *badmaash*; '4' Item highly associated with *badmaash*.

[2]**Importance** was rated on a 1 to 4 scale: '1' Item unlikely to lead to a "dark future"; '4' Item highly likely to lead to a dark future.