

Culturally Appropriate Instrument: A Systematic Literature Review

Herlina Siwi Widiana
Faculty of Psychology, Universitas Ahmad Dahlan
herlina.widiana@psy.uad.ac.id

Abstract

Culture has an impact on the presentation of mental health problems. Variation in symptoms, the existence of local idioms, and etiology were found in different target groups. Therefore, a culturally appropriate instrument is needed to assess mental health problems for a specific target group. This study aimed to explore the literature on developing a culturally appropriate instrument. A systematic literature review was conducted through the Web of Science Core Collection database search engine. The keywords applied for searching were depression, mental health, screening tool, scale, instrument, measure, assess, and culture. Only articles published in English were selected, which resulted in 8,113 articles. Refining the results was conducted based on several categories, resulting in 4,351 articles. The titles, abstracts, and main texts were then read. As a result, a final set of 33 articles was selected from a comprehensive review. A culturally appropriate instrument was developed by deriving items from experience and authoritative knowledge and then validating psychometric properties. Local idioms, symptoms, and constructs were colored the culturally appropriate instruments for specific target groups.

Keywords: culturally appropriate, mental health, psychological instrument, systematic literature review.

Received 12 September 2021/Accepted 29 October 2021 ©Author all rights reserved

Introduction

Considering both the content and context is essential to understand how mental health problems are constructed and present in different settings. Nearly four decades ago, Kleinman (1980) developed a framework for the explanatory model of illness, including etiology, causes, and treatment. This framework has been applied to understand depression in different cultural contexts. Previous studies have shown that local idioms of “thinking too much” as well as feelings of sadness are characteristic of depression. However, depression among Afghan refugees in the US was characterized by changes in temperament, altered cognitions, avoidance and dissociative behaviors, and somatic complaints (Alemi, Weller, Montgomery, & James, 2017). Somatic complaints have also been reported by people with depression in Denmark (Buus, Johannessen, & Stage, 2012). At the same time, Turkish women in Iran reported headaches and musculoskeletal pain (Dejman et al., 2011).

In addition to differences in manifestations, differences also occur in the diagnostic process. Research conducted among psychiatrists and GPs in Denmark shows that these two groups have different opinions about what depression is, with psychiatrists arguing that depression is a mental illness, while GPs argue that depression is a reaction to lived problems (Davidsen & Fosgerau, 2014). Such variations in etiology and the relative significance of different contextual factors result in different approaches to diagnosis. Psychiatrists used a direct approach, relying on clinical impressions and determining the severity based on instrumentation developed in psychiatry (such as Hamilton Depression Rating Scale), while GPs focused on the everyday problems of their patients (Davidsen & Fosgerau, 2014). Importantly, while psychiatrists questioned neither the diagnostic instrument nor the diagnosis, GPs thought that depression was a pragmatic construct, and they were generally critical of the value and accuracy of the instruments, diagnoses, guidelines, and rating scales created in psychiatry. They did not consider these instruments useful in primary health care settings and noted they had not been validated in the community. However, they also felt under pressure to use the instruments and make diagnosis consistent with these, even if they also felt that the diagnosis was misleading and that the approach ignored the complexities of people's experience which might have influence their mood.

Notwithstanding the significance of context, as described above in the Danish study, assessment is crucial in mental health, and accurate assessment is needed for accurate diagnosis that will lead to accurate treatment and appropriate care (McGorry & Os, 2013). The WHO (2017) has stated that one of the barriers to managing depression is inaccurate assessment, leading to false positive and false negative diagnoses, resulting in inappropriate treatment. Taking into consideration differences in the presentation and perceived etiology of depression across cultures, many researchers have argued the importance of a culturally appropriate screening tool, one that acknowledges a cultural overlay or underpinning, if and when that exists (Alang, 2016; Brintnell, Sommer, Kuncoro, Setiawan, & Bailey, 2013). As already noted, in the making this point 40 years ago, Kleinman (1987) recommended adding local idioms of depression instrumentation into standard questionnaires to accommodate cultural differences.

Developing a new instrument to measure a psychological construct can be done either with the "top down" or "bottom up" method. A theoretical-rational or deductive approach was

used in the top down method (Achenbach, Dumenci, & Rescorla, 2003). The first step is to conceptualize the construct, and then item pools are derived. The following steps are item and psychometric analysis (Clark & Watson, 1995). On the contrary, a bottom up method uses an inductive approach (Achenbach et al., 2003). As the bottom up method is empirically based, collecting data on particular behavior should be done first (McConaughy, 2001). Previous research on developing culturally appropriate tools used both different methods.

Previous systematic literature reviews were conducted in the mental health field. However, no systematic literature review was carried out on developing a culturally appropriate tool to screen mental health problems. Therefore, this study aimed to explore the existing literature on the development of a culturally appropriate tool. A compare and contrast approach were used to set the basis for developing a culturally sensitive screening tool with relevant psychometric properties.

Method

The PRISMA model was applied in a literature search. In identifying the relevant literature on developing local scales, the Web of Science Core Collection database search engine was used to conduct a search using TOPIC = (depress* OR “mental health”) AND TOPIC = (“screening tool” OR scale OR instrument OR measure* OR assess*) AND TOPIC = (cultur*). Only articles published in English were selected. This search method resulted in 8,113 articles. The result was then refined based on the following categories: psychiatry, psychology clinical, psychology multidisciplinary, psychology social, psychology, health care sciences services, social science biomedical, psychology applied, social sciences interdisciplinary, anthropology, ethnic studies, behavioral sciences, religion, social issues, and primary health care, with the document types including articles, book chapters, and reviews. These refinements resulted in 4,351 articles. After reading the titles, abstracts, and main texts, a final set of 33 articles was selected from a comprehensive review, as discussed below. Figure 1 illustrates the inclusion process.

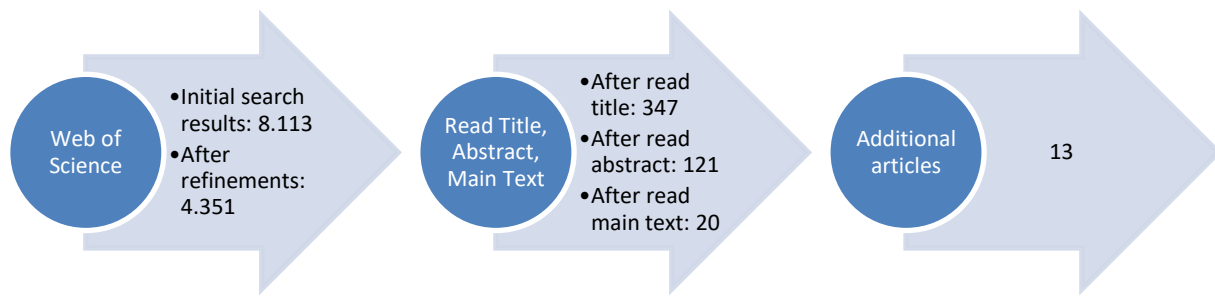


Figure 1. Process of inclusion (final number of articles: 33)

The final selection of articles for analysis was based on the main text. For this, 82 articles were excluded, as they did not relate to the development of culturally appropriate tools but rather included assessing depression in a specific population using “Western” standard assessment tools. Several articles reported psychometric properties of culturally appropriate assessment tools and the validation process. However, the absence of information on deriving items and developing the tool led to the exclusion of these articles. Nineteen articles explained the development of culturally appropriate tools, but these were not in the field of depression nor any other area of mental health.

By focusing on articles on developing a culturally appropriate tool to assess mental health problems, especially depression, 20 articles were identified in the Web of Science search. Thirteen additional articles focused on developing a culturally appropriate tool for mental health problems, particularly depression, were identified from searching through websites or directly contacting the authors. These articles were also reviewed, described, and analyzed.

Result

As explained above, 33 articles concerned with developing culturally appropriate tools were included in the analysis in the final selection. These articles were concerned with measuring depression (13), distress (8), stress (4), tension (2), nervousness among Latino adolescents in New York (1), mental health in general (3), and subjective wellbeing (2). Table 1 presents the sources of knowledge in developing the new tool. Experiences elicited from lay people, clinical participants, and experts were sources of knowledge to develop tools in 66.67% of the studies, while authoritative knowledge gained from the published literature, Western screening tools, or the DSM was used to develop new tools in 27.27% of the studies. Two

studies used both authoritative knowledge and experience as the source of information on which to base decisions relating to new tools.

Table I
Characteristics and the Source of Knowledge in Developing the Tool

Authors	Year	Construct	Knowledge	
			Authoritative	Experience
Kinzie et al.	1982	Depression		Experience of Vietnamese mental health workers
Zheng & Lin.	1991	Depression	Western depression screening tools & Chinese psychiatrists' judgment	
Cervantes, Padilla, & de Snyder.	1991	Stress		Experiences of lay people & judgment from experts
Abas, Phillips, Richards, Carter, & Levy.	1996	Emotional distress		Experiences of lay people & elderly clients
ElRufaie, Absood, & Abou-Saleh.	1997	Depression	Western depression screening tools	
Patel, Simunyu, Gwanzura, Lewis, & Mann.	1997	Idiom of distress		Experiences of patients & clinicians
Masse et al.	1998	Psychological distress (including depression)		Experiences of lay people
Kim.	2002	Depression		Experiences of lay people
Woo et al.	2004	Depression	Literature review & feedback from lay people	
Phan, Steel, & Silove.	2004	Psychiatric distress (depression, anxiety, somatization)	Literature review on Vietnamese traditional medicines, folk-medicines, novels, folktales, fairy tales, poetry, reviews of traditions, cultures, customs, and religions	An ethnographic survey among Vietnamese immigrants & Vietnamese clinical samples
Mumford et al.	2005	Anxiety & Depression		Psychiatrists case notes
Hung, Weng, Su, & Liu.	2006	Depression & somatic symptoms	DSM-IV & Western depression screening tools	
Miller et al.	2006	General Psychological Distress		Experiences of lay people
Koh, Chang, Fung, & Kee.	2007	Depression	Literature review & feedback from mental health professionals & target group	
Snipes, Thompson, O'Connor, Godina, & Ibarra.	2007	Stress		Experiences of lay people
Kaaya, Lee, Mbwambo, Smith-Fawzi, & Leshabari.	2008	Depression		Experiences of women who were suffering or had suffered from any of the local

				categories of illness, traditional healers, village health workers, and village leaders
Fernando G. A.	2008	Mental Health & Psychological Status		Experiences of lay people who are suffering as the impact of the tsunami
Betancourt et al.	2009	Mental health (depression, anxiety, conduct problem)		Opinion & experiences of lay people
Thomas, Cairney, Gunthorpe, Paradies, & Sayers.	2010	Subjective Wellbeing (including depression)	Literature review on indigenous Australia & consultation with indigenous people & indigenous mental health workers.	
Livanis & Tryon,	2010	Nervios	Literature review on nervios	
Weaver & Hadley.	2011	Tension		Perception of women with diabetic
Rasmussen, Katoni, Keller, & Wilkinson.	2011	Idiom of distress		Experiences of lay people & traditional healers
Praditsathaporn, Chandanasotthi, Amnatsatsuee, Nityasudd, & Sunsern.	2011	Mental health		Opinion of older lay people
Hwang et al.	2012	Depression	Western depression screening tools	Experiences of people with depression
Wong, Wu, Guo, Lam, & Snowden.	2012	Depression		Experiences of clinical participants & clinicians
	2012	Stress		Experiences of lay people & experts in Hispanic
Cervantes, Fisher, Cordova, & Napper.				
Cervantes, Goldbach, & Padilla.	2012	Stress		Experiences of lay people & experts in Hispanic mental health
				Opinion of lay people
Karasz, Patel, Kabita, & Shimu.	2013	Tension		
Kaiser, Kohrt, Keys, Khoury, & Brewster.	2013	Idiom of distress		Opinion & experiences of lay people, community leaders, & health workers
Rasmussen et al.	2015	Depression	Translation of Western depression screening tools & literature review of previous research among Haitian	
Xie et al.	2015	Depression	Western depression screening tools	
Mutumba et al.	2015	Psychological distress (including depression)		Experiences of adolescents living with HIV
Janca et al.	2015	Subjective & emotional wellbeing (including depression)		Opinion of Aboriginal stakeholders and mental health professionals.

Study population and sampling strategies

Random sampling is recommended in selecting a population sample for research, particularly where the method is quantitative, because of the reduced sample bias (Dattalo, 2010). However, only six studies applied this in their research, in both the deriving and validating stages (Patel et al., 1997; Phan et al., 2004; Praditsathaporn et al., 2011) or only in the validating stage (e.g., Fernando, 2008; Rasmussen et al., 2011; Woo et al., 2004). Other studies used convenience sampling, purposive sampling, or snowballing sampling, especially in the first stage when a qualitative approach was used to derive items for the new tool. This choice is understandable given that the new tool was being developed for a specific population. Therefore, the approach in stage one took advantage of qualitative methods to identify salient categories and experiences. In any given study, two different sampling methods might be applied, with purposive sampling in the qualitative stage to derive potential items and random or other sampling methods in the quantitative stage to evaluate psychometric properties of the new tool. Rasmussen et al. (2015), for example, also used two different sampling techniques in their work on developing a culturally appropriate depression screening tool, with purposive sampling used in the deriving stage and convenience sampling in the validating stage.

Across the articles, different considerations were used in assigning samples at each stage. In the first stage, where a qualitative approach was employed, several methods were applied, such as interviews, focus group discussions, free listing, surveys, observations, or case notes. Some studies used only one method; other studies combined two or more methods. Interviews were used in almost half of the studies (15 of 33), with the sample size ranging from 16 to 195 people. A small sample size was used with clinical case samples, while larger sample sizes were applied with lay people as research participants (e.g., Cervantes et al., 1991; Masse et al., 1998; Patel et al., 1997). At the validating stage, a larger sample size should be involved to enable appropriate statistical analysis. Some statistical analysis requires a minimum sample size, for example, with the use of structural equation modelling to conduct confirmatory factor analysis, when a minimum of 100 respondents is needed (Hair, Black, Babin, & Anderson, 2010). However, Mundfrom, Shaw, & Ke (2005) have argued that a small sample size in conducting factor analysis is influenced by the level of communality indicating the amount of variance and by the number of variables to be measured in each factor. The

majority of studies (69.7%) used a sample size of between 100 to 1000 participants in the validating stage. Only a few studies (12.12%) involved less than 100 participants (e.g., Janca et al., 2015; Kinzie et al., 1982; Snipes et al., 2007; Weaver & Hadley, 2011), and only three studies involved more than 1000 participants (e.g., Cervantes, Fisher, et al., 2012; Hwang et al., 2012; Praditsathaporn et al., 2011), by combining a clinical sample and community members as participants. Even though a larger sample size may lead to better precision (Biau, Kerneis, & Porcher, 2008), the sample size should be appropriate, neither too large nor too small (Faber & Fonseca, 2014). A critical consideration in determining the sample size is representativeness (Kaplan, Chambers, & Glasgow, 2014), with the primary consideration of sample representativeness in light of the study's objectives.

Most identified studies were conducted in the US and in Asia, and the Middle East (nine and eleven studies, respectively). Studies conducted in the US focused on immigrant populations and included studies with Bangladeshi (Karasz et al., 2013), Chinese (Wong et al., 2012), Mexican (Cervantes, Fisher, et al., 2012; Cervantes, Goldbach, et al., 2012; Cervantes et al., 1991; Livanis & Tryon, 2010; Snipes et al., 2007), Korean (Kim, 2002), and Vietnamese (Kinzie et al., 1982) immigrants. Studies conducted in Asian countries and the Middle East included Afghanistan (Miller et al., 2006), China (Xie et al., 2015; Zheng & Lin, 1991), India (Weaver & Hadley, 2011), Korea (Hwang et al., 2012), Pakistan (Mumford et al., 2005), Singapore (Koh et al., 2007; Woo et al., 2004), Sri Lanka (Fernando, 2008), Thailand (Praditsathaporn et al., 2011), and United Arab Emirates (ElRufaie et al., 1997). In total, 66.67% of studies were conducted with local majority populations, and 33.33% studies with immigrants in the US, UK, and Australia.

Each new tool reported in the articles has a specific ethnicity as the target population, the majority of Asian origin (51.52%), both in their home countries and as immigrants (12 and 5 studies, respectively). This result aligns with Kleinman's & Good argument (1985) that depression among Asians varies in presentation with their counterparts in Western societies. Nine studies were conducted among people of American origin, with most of the new tools developed for Mexican and Caribbean populations.

Most of the new tools reported in the articles were developed for adults, followed by adolescents and youth (66.67% and 18.18%, respectively). Three screening tools were

developed for older people, including older Caribbean populations (Abas et al., 1996), older Chinese (Xie et al., 2015), and older Thai (Praditsathaporn et al., 2011). Only one article reported developing a culturally appropriate depression screening tool for children (Koh et al., 2007). This distribution is in line with previous studies that depression can occur throughout the life span, although the highest prevalence is among people of working age (Ferrari et al., 2013). Most of the new tools were developed for both men and women, although one depression screening tool was developed especially for pregnant women in Tanzania (Kaaya et al., 2008).

Methods used in developing the tools

Studies on developing a culturally appropriate tool in the mental health field, especially relating to depression, have increased over the decades. The first article identified was published in 1982, measuring depression among Vietnamese immigrants in the US (Kinzie et al., 1982). Over the next decade, six articles were published, measuring broader psychological constructs, including depression (2), distress (3), and stress (1). The target population of these culturally specific tools also expanded, not only for immigrants but also for specific populations in their land of origin: Chinese in China (Zheng & Lin, 1991), Arabs in the United Arab Emirates (ElRufaie et al., 1997), Africans in Zimbabwe (Patel et al., 1997), and Québécois (French Canadians) in Canada (Masse et al., 1998). Thirteen articles were published from 2001 to 2010, with most of the tools developed associated with measuring depression (10 out of 13). In these articles, the tools measured depression as a single construct (Kaaya et al., 2008; Kim, 2002; Koh et al., 2007; Woo et al., 2004), which presented with other constructs: somatic symptoms (Hung et al., 2006), anxiety (Mumford et al., 2005), anxiety and somatization (Phan et al., 2004), depression as a part of another construct such as distress (Miller et al., 2006), (poor) mental health (Betancourt et al., 2009), or in association with subjective well-being (Thomas et al., 2010). A further thirteen articles were published between 2011 and May 2017. Only six tools measured depression, either as a single construct (Hwang et al., 2012; Rasmussen et al., 2015; Wong et al., 2012; Xie et al., 2015) or as part of another construct, including psychological distress (Mutumba et al., 2015) and subjective well-being (Janca et al., 2015).

In developing a new culturally appropriate mental health instrument, most studies combined a qualitative approach in deriving items and quantitative analysis in the validation stage

(69.70%). Interviews and focus group discussions were mainly used in the qualitative stage. While psychometric evaluations on reliability and validity were conducted in the quantitative stage.

Three different approaches were applied in deriving items. The first approach relied on the experience of patients (21.21%), while the second approach was based on the opinions of lay people or experts (45.45%). These two approaches were based on Kleinman's (1980) explanatory model of illness (EMs). The third approach selected items from previous standard assessment tools or a literature review, with judgment from experts on items relevant to specific populations (27.27%) that accommodate Kleinman's (1987) idea to add local idioms.

Most studies presented factor structure, reliability, and validity in validating stage. In analyzing factor structure, explanatory factor analysis (EFA) was applied in more studies than confirmatory factor analysis (CFA), 42.42% and 6.06%, respectively. 15.15% of these studies applied both EFA and CFA in series. A model of factor structure in CFA was developed based on the result of EFA (Fabrigar et al., 1999). Therefore, the factor structure model could be developed based on a theoretical or empirical basis (Fabrigar et al., 1999; Henson & Roberts, 2006).

Internal consistency was most frequently applied to estimate reliability (84.85%), as this method requires less effort (Furr & Bacharach, 2014; Streiner, 2003). Moreover, internal consistency seems to be more accurate in measuring mood, as mood changes over time (McCrae, Kurtz, Yamagata, & Terracciano, 2011). Only seven studies reported using both internal consistency and test-retest reliability, while none reported the parallel forms method.

Criterion related validity was frequent often reported in the studies (63.64%), followed by construct validity and content validity (30.30% and 21.21%, respectively). Construct validity depends on the content of the test, its internal structure, the psychological process used in test response, the association among test scores and other variables, and the consequences of test use (Furr & Bacharach, 2014). Therefore, content validity, factor structure, and criterion validity influence construct validity, so validity relies on more than one statistic

(Coaley, 2010).

Instrumentation

Not all authors clearly stated how the new tool was administered (e.g., Hwang et al., 2012; Zheng & Lin, 1991). The identification of how to administer the tool through self-administration or interview was determined from explanations in the text. In nearly half of the reviewed articles, the tools were administered through self-report only. Several terms were used in describing this, including self-report, self-inventory, and self-administration. However, other articles reported that the new tools developed were administered by interview only or by both interview and self-report (33.33% and 12.12%, respectively). It can be said that nearly half of the culturally appropriate tools were delivered through interviews. One of the reasons why the tools were administered verbally is low literacy, such as among indigenous youth in Australia (Thomas et al., 2010), Arabic people attending primary health centers (ElRufaie et al., 1997), and Acholi youth in Northern Uganda (Betancourt et al., 2009). In this regard, the characteristics of target populations influenced the format of the tool. In this context, awareness is needed in developing a culturally appropriate tool, ensuring that both the content and design reflect the specific target population.

Based on explanations in the text and numbers included in tables, the length of the final instruments can be determined. The length of the final scale of the majority of studies was around 20 items. The percentage of studies with between 21 -35 final items and 20 items or less were nearly the same (42.42% and 39.39%, respectively). No specific number could be referred to as an ideal length of a scale. The advantage of a shorter scale is that it is less time-consuming (Morgado, Meireles, Neves, Amaral, & Ferreira, 2017). However, the length of the scale is associated with reliability; the larger the number of items in a scale, the greater the scale's reliability (DeVellis, 2003). Therefore, reliability should be considered in determining the length of a scale, particularly when subscales are included, with an ideal Alpha ranging from .8 to .9 (Morgado et al., 2017) and acceptable above .7 (Hills, 2011).

Specific symptoms in the instrument

Differences in symptoms, the local idioms, and terminology were identified in these studies. Different symptoms were identified among older Caribbean residents in London, UK (Abas et al., 1996), Acholi youth in Northern Uganda (Betancourt et al., 2009), and Vietnamese

(Kinzie et al., 1982). Depression was also presented in a social context, including among Pakistanis in Peshawar and Lahore (Mumford et al., 2005), Singaporean Chinese children (Koh et al., 2007), and Asian adolescents in Singapore (Woo et al., 2004). Therefore, social items were added in the culturally appropriate depression screening tool for Chinese American immigrants (Wong et al., 2012).

Somatic symptoms were presented in the Kim Depression Scale for Korean American (KDSKA) (Kim, 2002), the Lee and Rhee Depression Scale (LRDS), the Dar-es-Salaam Symptom Questionnaire (DSQ) (Kaaya et al., 2008), and the Phan Vietnamese Psychiatric Scale (PVPS) (Phan et al., 2004). Some studies reported externalized emotions as the manifestation of depression, for example, among indigenous Australians (Thomas et al., 2010) and Chinese American immigrants in the US (Wong et al., 2012). Religiosity was also found in the instruments reviewed, including in the Pakistan Anxiety and Depression Questionnaire (PADQ) (Mumford et al., 2005) and the Chinese American Depression Scale (CADS) (Wong et al., 2012).

Local idioms of distress were also found in the Afghan Symptom Checklist (ASC) (Miller et al., 2006), a psychological distress scale for ALH in Uganda (Mutumba et al., 2015), and Zanmi Lasante Depression Symptom Inventory (ZLDSI) (Rasmussen et al., 2015), Kreyol Distress Idiom (KDI) (Kaiser et al., 2013), the CADS (Wong et al., 2012), the Sri Lankan Index of Psychosocial Status--Adult Version (SLIPSS-A) (Fernando, 2008), the Shona Symptom Questionnaire (SSQ) (Patel et al., 1997). Moreover, different constructs on depression have been identified among Darfuris (Rasmussen et al., 2011), Indian women (Weaver & Hadley, 2011), and Québécois (Masse et al., 1998). Therefore, an emic approach is needed in assessment as a cultural overlay requires inclusion.

Discussion

The growing number of research articles on developing a culturally appropriate tool in mental health fields, including depression, highlights an understanding that the presentation of mental health problems varies across cultural backgrounds. Therefore, to assess depression accurately, an exploration of local presentation is needed. The main source of knowledge in developing a culturally appropriate tool among articles reviewed was based on either experience of patients or opinion from experts and lay people (66.7%) and indicated the

importance of an emic approach as the basis of its development.

The explanatory model of illness introduced by Kleinman (1980) is one way to gain a cultural understanding of depression among a target population, and this is based on the perception of people with depression concerning: their illness, its cause, how they present it, how they seek treatment and from whom they seek treatment. As explained above, an explanatory model of illness was included in the Cultural Formulation Interview (CFI) of the newest DSM, the DSM-5 (APA, 2013; Mezzich, Caracci, Fabrega Jr, & Kirmayer, 2009). The CFI was developed to improve psychiatric diagnostic and consider cultural factors (Worcester, 2013), explore local presentation, and better understand mental health problems, including depression among different communities.

As described above, most studies (69.70%) used a combination of qualitative and quantitative methods, whereby a qualitative approach was used in the initial stages to derive candidate items from being tested in a draft instrument. A qualitative approach is important when trying to identify the distinctive experiences and manifestations of depression across cultural backgrounds, with data generated through interviews, focus group discussions, free listing, surveys, observations, or case notes. The design of a culturally appropriate tool for depression in a specific target group focuses on understanding how people speak about and present their depression, and qualitative methods seem the best approach to achieve this goal (Sullivan & Sargeant, 2011). By applying a qualitative approach, detailed information from the experience of individuals will be gained (Stewart, Makwarimba, Barnfather, Letourneau, & Neufeld, 2008).

In the qualitative phase of the articles reviewed, the authors explored how target populations presented their mental health problems. As described above, specific symptoms were identified. Some symptoms were identified in some populations but through the use of different local idioms, such as thinking too much. This specific presentation indicates the importance of local language in communicating the symptoms, that is, what patients experience and feel. To ensure the newly developed tool is culturally sensitive, these particular symptoms need to be included in an assessment tool targeting a given population, as Kleinman (1987) suggested.

Crossover analyses in the mixed method framework allow researchers to switch between a qualitative and quantitative lens (Onwuegbuzie, Bustamante, & Nelson, 2010). By applying crossover analysis in developing a new instrument with a mixed-method design, there is the possibility to blend a qualitative approach in the initial stage when deriving items for the new tool with quantitative analysis in the validation stage, included in factor analysis (Onwuegbuzie, 2003). Here, the results of the quantitative analysis evaluate whether emergent themes derived in the initial qualitative method are supported by data gathered from a larger sample in the given target population. This method provides support for the implementation of the newly developed tool within the target population.

Even though universality in the presentation of depression exists, in manuals, for example, the DSM, this “objective” guidance is based on an etic approach. However, differences in details of depression have been documented, and an emic approach is needed to fully understand depression across cultures to appropriately assess and manage depression.

There is a strong argument that a culturally sensitive depression assessment tool is required to assess depression in different populations. To develop this, the combination of a qualitative approach in deriving items and quantitative analysis in validating the new tool appears the most appropriate. Therefore, local idioms and specific experiences gained through an explanatory model are required. A qualitative method may work better in the initial stages of developing a tool to gain a local understanding and experience around the nuances of the presentation of depression. Subsequently, a quantitative method assessing the psychometric properties of the newly developed tool will enhance the scale’s construct validity. Quantitative analysis with some statistical indicators is important as the basis of implementing the culturally appropriate tool in clinical practice.

Conclusion

The first culturally appropriate instrument was developed by Kinzie et al. in 1982 to measure depression among Vietnamese immigrants in the US. Literature reported majority target group of the instruments were Asian adults. The culturally appropriate instruments were developed in the form of self-inventory rather than delivered with the interview. Two stages, deriving items and validating the instrument, were applied to develop the culturally appropriate instrument, with various sampling techniques and sample sizes. The experience

was used more than authoritative knowledge in deriving items, with local idioms, terminology, and symptoms in the instruments developed to address the cultural variations in a specific target group.

References

- Abas, M., Phillips, C., Richards, M., Carter, J., & Levy, R. (1996). Initial development of a Cnew culture-specific screen for emotional distress in older Caribbean people. *International Journal of Geriatric Psychiatry, 11*(12), 1097–1103.
- Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32*(3), 328–340.
- Alang, S. M. (2016). “Black folk don’t get no severe depression”: Meanings and expressions of depression in a predominantly black urban neighborhood in Midwestern United States. *Social Science & Medicine, 157*(1–8).
- Alemi, Q., Weller, S. C., Montgomery, S., & James, S. (2017). Afghan refugee explanatory models of depression: Exploring core cultural beliefs and gender variations. *Medical Anthropology Quarterly, 31*(2), 177–197.
- APA. (2013). *American Psychiatric Association: Diagnostic and Statistical Manual on Mental Disorder* (Fifth ed.). Arlington, VA: American Psychiatric Association.
- Betancourt, T. S., Bass, J., Borisova, I., Neugebauer, R., Speelman, L., Onyango, G., & Bolton, P. (2009). Assessing local instrument reliability and validity: A field-based example from northern Uganda. *Social Psychiatry and Psychiatric Epidemiology, 44*(8), 685–692.
- Biau, D. J., Kerneis, S., & Porcher, R. (2008). Statistics in brief: The importance of sample size in the planning and interpretation of medical research. *Clinical Orthopaedics and Related Research, 466*(9), 2282–2288.
- Brintnell, E. S., Sommer, R. W., Kuncoro, B., Setiawan, G. P., & Bailey, P. (2013). The expression of depression among Javanese patients with major depressive disorder: A concept mapping study. *Transcultural Psychiatry, 50*(4), 579–598.
- Buus, N., Johannessen, H., & Stage, K. B. (2012). Explanatory models of depression and treatment adherence to antidepressant medication: A qualitative interview study. *International Journal of Nursing Studies, 49*(10), 1220–1229.
- Cervantes, R. C., Fisher, D. G., Cordova, D., & Napper, L. E. (2012). The Hispanic Stress Inventory-Adolescent version: A culturally informed psychosocial assessment. *Psychological Assessment, 24*(1), 187–196.
- Cervantes, R. C., Goldbach, J. T., & Padilla, A. M. (2012). Using qualitative methods for revising items in the Hispanic Stress Inventory. *Hispanic Journal of Behavioral Sciences, 34*(2), 208–231.

- Cervantes, R. C., Padilla, A. M., & de Snyder, N. S. (1991). The Hispanic Stress Inventory: A culturally relevant approach. *Psychological Assessment*, 3(3), 001–010.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319.
- Coaley, K. (2010). *An introduction to psychological assessment and psychometrics*. London: SAGE Publication, Ltd.
- Dattalo, P. (2010). *Strategies to approximate random sampling and assignment*. New York: Oxford University Press.
- Davidson, A. S., & Fosgerau, C. F. (2014). What is depression? Psychiatrists' and GPs' experiences of diagnosis and the diagnostic process. *International Journal of Qualitative Studies on Health and Well-Being*, 9.
- Dejman, M., Forouzan, A. S., Assari, S., Malekafzali, H., Nohesara, S., Khatibzadeh, N., ... Ekblad, S. (2011). An explanatory model of depression among female patients in Fars, Kurds, Turks ethnic groups of Iran. *Iranian Journal of Public Health*, 40(3), 79–88.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2 ed.). Thousand Oaks, California: Sage Publications, Inc.
- EIRufaie, O. E. F., Absood, G. H., & Abou-Saleh, M. T. (1997). The Primary Care Anxiety and Depression (PCAD) Scale: A culture-oriented screening scale. *Acta Psychiatrica Scandinavica*, 95(2), 119–124.
- Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics*, 19(4), 27–29.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
- Fernando, G. A. (2008). Assessing mental health and psychosocial status in communities exposed to traumatic events: Sri Lanka as an Example. *American Journal of Orthopsychiatry*, 78(2), 229–239.
- Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., ... Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, 10(11).
- Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction* (Second ed.). Thousand Oaks, CA: SAGE Publication, Inc.
- Hair, J. F., Black, W., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (Seventh ed.). New Jersey: Pearson Prentice Hall.

- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research - Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66(3), 393–416.
- Hills, A. M. (2011). *Foolproof guide to statistic using IBM SPSS* (Second ed.). Frenchs Forest, NSW: Pearson Australia.
- Hung, C. I., Weng, L. J., Su, Y. J., & Liu, C. Y. (2006). Preliminary study of a scale measuring depression and somatic symptoms. *Psychological Reports*, 99(2), 379–389.
- Hwang, S. H., Rhee, M. K., Kang, R. H., Lee, H. Y., Ham, B. J., Lee, Y. S., & Lee, M. S. (2012). Development and validation of a screening scale for depression in Korea: The Lee and Rhee Depression Scale. *Psychiatry Investigation*, 9(1), 36–44.
- Janca, A., Lyons, Z., Balaratnasingam, S., Parfitt, D., Davison, S., & Laugharne, J. (2015). Here and Now Aboriginal Assessment: background, development and preliminary evaluation of a culturally appropriate screening tool. *Australasian Psychiatry*, 23(3), 287–292.
- Kaaya, S. F., Lee, B., Mbwambo, J. K., Smith-Fawzi, M. C., & Leshabari, M. T. (2008). Detecting depressive disorder with a 19-item local instrument in Tanzania. *International Journal of Social Psychiatry*, 54(1), 21–33.
- Kaiser, B. N., Kohrt, B. A., Keys, H. M., Houry, N. M., & Brewster, A. R. T. (2013). Strategies for assessing mental health in Haiti: Local instrument development and transcultural translation. *Transcultural Psychiatry*, 50(4), 532–558.
- Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014). Big data and large sample size: A cautionary note on the potential for bias. *Cts-Clinical and Translational Science*, 7(4), 342–346.
- Karasz, A., Patel, V., Kabita, M., & Shimu, P. (2013). “Tension” in South Asian women: Developing a measure of common mental disorder using participatory methods. *Progress in Community Health Partnerships-Research Education and Action*, 7(4), 429–441.
- Kim, M. T. (2002). Measuring depression in Korean Americans: Development of the Kim Depression Scale for Korean Americans. *Journal of Transcultural Nursing*, 13(2), 109–117.
- Kinzie, J. D., Manson, S. M., Vinh, D. T., Tolan, N. T., Anh, B., & Pho, T. N. (1982). Development and validation of a Vietnamese-language depression rating-scale. *American Journal of Psychiatry*, 139(10), 1276–1281.
- Kleinman, A. (1980). *Patients and healers in the context of culture: An exploration on the borderland between Anthropology, Medicine and Psychiatry*. Berkeley: University of California Press.
- Kleinman, A. (1987). Anthropology and psychiatry - The role of culture in cross-cultural research on illness. *British Journal of Psychiatry*, 151, 447–454.
- Kleinman, A., & Good, B. (1985). *Culture and depression*. California: University of California Press.

- Koh, J. B. K., Chang, W. N. C., Fung, D. S. S., & Kee, C. H. Y. (2007). Conceptualization and manifestation of depression in an Asian context: Formal construction and validation of a Children's Depression Scale in Singapore. *Culture Medicine and Psychiatry*, 31(2), 225–249.
- Livanis, A., & Tryon, G. S. (2010). The development of the Adolescent Nervios Scale: Preliminary findings. *Cultural Diversity & Ethnic Minority Psychology*, 16(1), 9–15.
- Masse, R., Poulin, C., Dassa, C., Lambert, J., Belair, S., & Battaglini, A. (1998). The structure of mental health: Higher-order confirmatory factor analyses of psychological distress and well-being measures. *Social Indicators Research*, 45(1–3), 475–504.
- McConaughy, S. H. (2001). The Achenbach system of empirically based assessment. In G. D. Pbye, D. H. Saklofske, J. J. W. Andrews & H. L. Janzen (Eds.). *Handbook of Psychoeducational assessment a practical handbook* (pp. 289–234). Burlington: Elsevier Science.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for Personality Scale Validity. *Personality and Social Psychology Review*, 15(1), 28–50.
- McGorry, P., & Os, J. V. (2013). Redeeming diagnosis in psychiatry: Timing versus specificity. *The Lancet*, 381, 343–345.
- Mezrich, J. E., Caracci, G., Fabrega Jr, H., & Kirmayer, L. J. (2009). Cultural formulation guidelines. *Transcultural Psychiatry*, 46, 383–405.
- Miller, K. E., Omidian, P., Quraishy, A. S., Quraishy, N., Nasiry, M. N., Nasiry, S., ... Yaqubi, A. (2006). The Afghan symptom checklist: A culturally grounded approach to mental health assessment in a conflict zone. *American Journal of Orthopsychiatry*, 76(4), 423–433.
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia-Reflexao E Critica*, 30(1), 1–20.
- Mumford, D. B., Ayub, M., Karim, R., Izhar, N., Asif, A., & Bavington, J. T. (2005). Development and validation of a questionnaire for anxiety and depression in Pakistan. *Journal of Affective Disorders*, 88(2), 175–182.
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168.
- Mutumba, M., Resnicow, K., Bauermeister, J. A., Harper, G. W., Musiime, V., Snow, R. C., & Lepkowski, J. M. (2015). Development of a psychosocial distress measure for Ugandan Adolescents living with HIV. *Aids and Behavior*, 19(2), 380–392.
- Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity*, 37(4), 393–409.

- Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4*(1), 56–78.
- Patel, V., Simunyu, E., Gwanzura, F., Lewis, G., & Mann, A. (1997). The Shona symptom questionnaire: The development of an indigenous measure of common mental disorders in Harare. *Acta Psychiatrica Scandinavica, 95*(6), 469–475.
- Phan, T., Steel, Z., & Silove, D. (2004). An ethnographically derived measure of anxiety, depression and somatization: The Phan Vietnamese Psychiatric Scale. *Transcultural Psychiatry, 41*(2), 200–232.
- Praditsathaporn, C., Chandanasotthi, P., Amnatsatsuee, K., Nityasudd, D., & Sunsern, R. (2011). Mental health assessment tool for older Thai adults: Development and psychometric testing. *Southeast Asian Journal of Tropical Medicine and Public Health, 42*(3), 744–753.
- Rasmussen, A., Eustache, E., Raviola, G., Kaiser, B., Grelotti, D. J., & Belkin, G. S. (2015). Development and validation of a Haitian Creole screening instrument for depression. *Transcultural Psychiatry, 52*(1), 33–57.
- Rasmussen, A., Katoni, B., Keller, A. S., & Wilkinson, J. (2011). Posttraumatic idioms of distress among Darfur refugees: Hozun and Majnun. *Transcultural Psychiatry, 48*(4), 392–415.
- Snipes, S. A., Thompson, B., O'Connor, K., Godina, R., & Ibarra, G. (2007). Anthropological and psychological merge: Design of a stress measure for Mexican farmworkers. *Culture Medicine and Psychiatry, 31*(3), 359–388.
- Stewart, M., Makwarimba, E., Barnfather, A., Letourneau, N., & Neufeld, A. (2008). Researching reducing health disparities: Mixed methods approach. *Social Science & Medicine, 66*(6), 1406–1417.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103.
- Sullivan, G. M., & Sargeant, J. (2011). Qualities of qualitative research: Part I. *Journal of Graduate Medical Education, 3*(4), 449–452.
- Thomas, A., Cairney, S., Gunthorpe, W., Paradies, Y., & Sayers, S. (2010). Strong souls: Development and validation of a culturally appropriate tool for assessment of social and emotional well-being in indigenous youth. *Australian and New Zealand Journal of Psychiatry, 44*(1), 40–48.
- Weaver, L. J., & Hadley, C. (2011). Social pathways in the comorbidity between type 2 diabetes and mental health concerns in a pilot study of urban middle- and upper-class Indian women. *Ethos, 39*, 211–225.
- WHO. (2017). Depression. <http://www.who.int/mediacentre/factsheets/fs369/en/>. accessed 1

May 2017.

- Wong, R., Wu, R., Guo, C., Lam, J. K., & Snowden, L. R. (2012). Culturally sensitive depression assessment for Chinese American immigrants: Development of a comprehensive measure and a screening scale using an item response approach. *Asian American Journal of Psychology, 3*(4), 230–253.
- Woo, B. S. C., Chang, W. C., Fung, D. S. S., Koh, J. B. K., Leong, J. S. F., Kee, C. H. Y., & Seah, C. K. F. (2004). Development and validation of a depression scale for Asian adolescents. *Journal of Adolescence, 2*, 677–689.
- Worcester, S. (2013). New interview enhances DSM-5 cultural formulation. *Family Practice News, 43*(9), 16.
- Xie, Z. J., Lv, X. Z., Hu, Y. D., Ma, W. X., Xie, H. G., Lin, K., ... Wang, H. L. (2015). Development and validation of the geriatric depression inventory in Chinese culture. *International Psychogeriatrics, 27*(9), 1505–1511.
- Zheng, Y. P., & Lin, K. M. (1991). Comparison of The Chinese Depression Inventory and the Chinese version of The Beck Depression Inventory. *Acta Psychiatrica Scandinavica, 84*(6), 531–536.