Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby gran to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	___24th May 2019___
Amna Hassana Abdulsalam	Date

Status of Mental Health in Sub-Saharan Africa: A Systematic Review of the Cultural Validity of Screening Tools

By

Amna Abdulsalam
Degree to be awarded: MPH

Hubert Department of Global Health

_____ [Chair's signature]
Jessica Fairley, MD, MPH
Committee Chair

STATUS OF MENTAL HEALTH IN SUB-SAHARAN AFRICA: A SYSTEMATIC REVIEW OF THE CULTURAL VALIDITY OF SCREENING TOOLS

By

Amna Hassana Abdulsalam

B.S., University of Wisconsin-Madison - 2016

THESIS COMMITTEE CHAIR:

Jessica K. Fairley, MD, MPH
Hubert Department of Global Health
Director, Emory Hansen's Disease Program
Associate Professor of Medicine
Division of Infectious Diseases / TravelWell Center

24th April 2019

An abstract of a
thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements
for the degree of
Master of Public Health
in Hubert Department of Global Health

Abstract

Status of Mental Health in Sub-Saharan Africa: A Systematic Review of the Cultural Validity of Screening Tools
By Amna Hassana Abdulsalam

Background: Mental health burden is on the rise in all countries across the world, especially in sub-Saharan Africa, as the region faces a growing and aging population, recent and on-going conflicts, economic difficulties and tens of millions of people living with HIV/AIDS. Despite this, mental health care in the region remains sub-optimal. Mental health screenings are a fundamental way to help bridge this gap and allow for early identification and intervention. They can be particularly useful is in sub-Saharan Africa for non-specialists to task shift and provide community-based care. However, one of the greatest challenges medical professionals in Africa are faced with is how to incorporate cross-cultural understandings of mental disorders into psychiatric screening of their patients. Literature reveals a wealth of studies on mental disorder screening tools but only a few that attempt to systematically review their validity in this setting. There is an emergent need to systematize and strengthen screening tools in sub-Saharan Africa. There remains a gap in knowledge of the current attempts at ensuring cultural validity in screening instruments, of all mental disorders, in sub-Saharan Africa. This review was conducted to identify the construct and cultural validity of mental disorder screening tools and how this impacts their reliability for use in these settings.

Methods: Four electronic databases (Pubmed, Web of Science, Popline and PsychINFO) were searched for published papers on cultural validity of mental health screening tools in sub-Saharan Africa. Original papers, published in English, on all ICD-10 mental disorders in sub-Saharan were included.

Results: Fifty-five studies satisfied the inclusion criteria. Overall, language translation (56%), adaptation of terms (47%) and method of administration (72%) were found to be the main methods of culturally validating screening tools. Back translation of tools into the local language was the most frequent translation method. Adaptation of terms was done through group consensus meetings by research experts and translators (46%), focus groups with community members (12%), pre-study pilot testing in a sample of the study population (8%) whilst the remainder of studies (37%) utilized a combination of the methods. 40 studies (72%) administered tools through interviews as opposed to using self-administered tools. Additionally, data analysis was the primary basis for concluding the validity of the instruments being measured in all the studies. 33 out of 55 studies (60%) developed a Receiver Operating Curve to calculate sensitivity and specificity.

Conclusion: Themes identified in this review highlight the strong methodological techniques utilized to cultural validity screening tools in sub-Saharan Africa, and the recommendations proposed should be used to stimulate meaningful discussions to improve these methods. This is essential to promote the best screening tools for use in sub-Saharan Africa and to relieve the rising burden of mental disorders in the region.

STATUS OF MENTAL HEALTH IN SUB-SAHARAN AFRICA: A SYSTEMATIC REVIEW OF THE CULTURAL VALIDITY OF SCREENING TOOLS

By

Amna Hassana Abdulsalam

B.S., University of Wisconsin-Madison - 2016

THESIS COMMITTEE CHAIR:

Jessica K. Fairley, MD, MPH
Hubert Department of Global Health
Director, Emory Hansen's Disease Program
Associate Professor of Medicine
Division of Infectious Diseases / TravelWell Center

24th April 2019

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree
of
Master of Public Health
in Hubert Department of Global Health

Acknowledgements

My most sincere thanks and appreciation go to the following:

First and foremost, to almighty God who helped me through every step of this MPH and thesis process and for without my faith I am nothing.

Dr. Jessica Fairley who supported my work and has provided excellent mentorship and support through this thesis process.

Dr. Elizabeth Walker who is an esteemed public mental health professional and whose class inspired this topic.

Flavia Traven for being the best and most supportive ADAP and providing me with unconditional support every step of this MPH process.

To all who reviewed this thesis and whose meticulous feedback helped get this project to completion: Nasir Zubairu, Maria Sundaram, Ahmed Jibril, Aisha Jibril and Hafsat Abdulsalam.

To Mom for unfailing and endless encouragement all my life and through this project; and Dad for always encouraging me to aim for the highest dreams and who instilled hard work and perseverance in me, and above all believed in me.

And finally, to all the MPH students who supported tirelessly by my side and provided the best moral support.

Table of Contents

I. Introduction

1.1 Importance of Mental Health

Mental disorders are defined as a combination of "abnormal thoughts, perceptions, emotions, behaviour and relationships with others" (World Health Organization [WHO], 2018). Mental disorders include the following: depression, bipolar affective disorder, schizophrenia and other psychoses, dementia, intellectual disabilities and developmental disorders including autism. They pose one of the greatest burdens to overall human health and lead to significant and chronic, health impacts, as well as extensive consequences on human rights and world economies (World Health Organization [WHO], 2018).

Mental health burden is on the rise in all countries across the world as many health systems are ill-equipped to sufficiently respond to mental disorders. Thus, a global gap exists between the mental health needs of populations and the delivery of treatment. In high-income countries, it is estimated that about 35% - 50% of those with mental disorders do not receive treatment whereas in low-and-middle-income countries, this number is estimated at 76% - 85% (World Health Organization [WHO], 2018). Another factor contributing to the increase in the burden of mental health is the poor quality of care delivered to those who do receive treatment. Many mental health policies and programs are based on outdated knowledge about mental disorders and failure to account for the differences in symptomatic presentation of disorders in different cultures. Additionally, even countries with more progressive legislation relating to mental health still face the issue of underfunding (World Health Organization, 2018). Consequently, this leaves primary care facilities, which are the most accessible settings to the largest number of people, lacking general health personnel who are adequately trained in mental health care and essential

psychotropic medications (World Health Organization, & UNAIDS, 2001). This particularly worsens mental disorders for many individuals as they delay treatment when it is not easily accessible. Research has shown that integrating care is crucial to addressing all the healthcare needs of individuals with mental health problems (Butler et al., 2008). Beyond receiving medical care, those with mental disorders also require social support and community care. Developments in social support are inhibited by the stigma surrounding mental disorders in many countries, which prevents access to education, employment, and housing for those with mental disorders. Additionally, fear of repercussions from this stigma discourages individuals from seeking care (World Health Organization [WHO], 2018). This is rooted in the belief that those with mental disorders are a danger to society and are likely to be violent towards others. Media reports exacerbate this unfounded belief by stereotyping those with mental illnesses. Stigma is cyclical to the mental illness as the social isolation and fear of treatment tends to worsen a person's mental health problems. This ultimately hinders people from actively participating and living in their communities ("Stigma and discrimination," 2015).

The above problems persist in countries all over the world, but progression and symptomology are greatly influenced by individual societies and cultures, as illustrated by differences in the mental health treatment gap in low versus high income countries. Lower-income countries, in particular, are experiencing a rise in mental disorders (World Health Organization [WHO], 2017). According to the World Health Organization (2017), there are now 29.19 million cases of depressive disorder in sub-Saharan Africa. This accounts for the second largest contributor to non-fatal health loss (7.9% of all years lived with disability [DALYs]) in the African region. Additionally, anxiety disorders affect approximately 25.91 million people and account for 2.9%

of all years lived with disability (DALYs). Despite this, mental disorders remain largely under-acknowledged in official health policies in sub-Saharan Africa and there remains a distinct lack of services and treatments due to political, economic, and institutional processes (Akyeampong, Hill & Kleinman, 2015).

One such process in Africa is a steadily growing population (Akyeampong, Hill & Kleinman, 2015). Life expectancy in all sub-Saharan African countries has been on the rise since the millennium. Specifically, no sub-Saharan African countries have experienced any decreases in life expectancy and rises have been as significant as 20% - 42% in some countries (Johnson, 2016). As such, more people are living to and beyond their mid-forties, the age when depression and anxiety most commonly occurs (World Health Organization [WHO], 2017). Other processes largely contributing to Africa's high mental health burden are the numerous post-colonial conflicts that have occurred and the rise of HIV/AIDS since the late 1980s (Akyeampong, Hill & Kleinman, 2015). Since the late 1950s, all 46 sub-Saharan African countries, as well as Somalia, Djibouti and Sudan, have been plagued by a multitude of conflicts due to efforts to gain independence and attempts to establish a stable post-colonial system. Furthermore, conflicts have not only arisen from active attempts at gaining independence but also from the governance structures established post-independence, as ethnic and political factions from pre-colonial years still existed. While there has been a substantial decline in large scale conflicts since 1990, the effects of war on individuals and societies still remains under-addressed in African societies and continues to contribute to chronic mental disorders in the entire region (Africa Conflict Prevention Pool (Great Britain) & Marshall, 2005). Finally, as mentioned above, HIV remains a major global health issue with significant implications for mental health. This burden is very

prominent in sub-Saharan Africa where 66% of the world's estimated 36.9 million people living with HIV, living in sub-Saharan Africa. Among these, 800,000 were new HIV infections, in 2017 ("Global HIV and AIDS Statistics", 2018). The association between HIV/AIDS and mental illness is complex and bidirectional. Although there is minimal research on this topic in sub-Saharan Africa, a review by Breuer, Myer, Struthers, & Joska (2011) on mental health and HIV/AIDS in sub-Saharan Africa highlights the high prevalence of mental illness among those living with HIV/AIDS. Depression was particularly prominent in this population and was worse for those at greater stages of their illness. Thus, the presence of chronic physical symptoms is evidently a factor in mental disorders. Additionally, due to misconceptions surrounding the methods of transmission of HIV, positive individuals face discrimination and social isolation which is a predisposing factor for many mental disorders (Breuer, Myer, Struthers, & Joska, 2011).

These three factors, though not sole contributors, are notable to the burden of mental health in sub-Saharan Africa. Addressing mental health in the sub-Saharan African region requires both a thorough understanding of the unique predisposing factors affecting individuals' and communities' mental health, as well as adoption of resources and practices that have worked best globally (World Health Organization, 2018).

1.2 Importance of Screening Tools

Early detection and treatment are crucial to reducing the burden of mental disorders as prevention of the adverse course of mental disorders is essential for the overall health of individuals. Although the symptoms of most chronic mental health conditions manifest in early to mid-adult years (24 -45 years), research shows that most disorders actually have early onset

years, with approximately 50% of chronic disorders beginning by the age 14 and 75% by the age 24 (NAMI, n.d.). The reason for this discrepancy is thought to be lack of adequate screening techniques and access to care for people with mental disorders, that is although onset occurs early, many people disregard their symptoms until they become unbearable (Mojtabai et al., 2011). Additionally, literature shows that the average time between onset of symptoms and intervention is approximately 10 years. Mental health screenings are a fundamental way to help bridge this gap and allow for early identification and intervention. Screening helps evaluate genetic, environmental and behavioural risk factors for disorders and distinguish between levels of intervention needed. Research has also shown early treatment results in both, better health outcomes and lessens the negative course of long-term disability (NAMI, n.d.). The U.S. Preventive Services Task Force (USPSTF) officially recommends that primary care physicians should regularly screen their adult patients for depression in clinical settings to ensure early detection of depression and to begin effective treatment and follow up. This recommendation is based on a review of numerous studies that illustrates the moderate net benefit of screening in adults as well as sub groups such as pregnant and postpartum women (O'Connor, Rossom, Henninger, Groom, & Burda, 2016). USPSTF states that screenings help detect symptoms that might otherwise go undetected (Siu et al., 2016).

Mental health screening for common disorders is on the rise in most primary care settings in high-income countries. However, there still remains barriers in these settings. Studies show that of "individuals who die by suicide, approximately 90% had a mental health disorder, and 40% of these individuals had visited their primary care doctor within the month before their suicide" (Miller, 2014). This serves as evidence of the lack of adequate screenings for mental health in

primary healthcare (Miller, 2014). This problem is exacerbated in low resource settings such as sub-Saharan Africa where there are even fewer health workers and even less mental health specialists. It is in these settings that screening tools can be crucial for non-specialists to task shift and provide community-based care (Sweetland, Belkin & Verdeli, 2013). For screening to accurately predict mental disorders, screening tools must be reliable, valid, sensitive and specific to the community and population they seek to aid. Validity in screening tests allows for the accurate identification of disorders, whereas reliability allows for the reproducibility of results across whole populations. A valid and reliable screening tool is therefore one that correctly identifies a disorder and does so correctly for everyone it is administered to in that population. Sensitivity measures the ability of the test to correctly identify disorders in individuals with a disorder whilst specificity measures the ability of the test in correctly identifying no disorder in individuals with no disorder. These are important characteristics for tools to ensure that they not only correctly identify disorders but also correctly identify and characterize non-disorders as such (American Academy of Pediatrics, 2012).

1.3 Importance of Cultural Validity

The high mental health burden in sub-Saharan Africa has to be addressed in order to advance economic status and improve morbidity and mortality rates. As highlighted above, screening tools can be an effective intervention to alleviate this burden, however, to succeed in this setting, Africa's history and current struggles with mental health must be carefully considered. One of the greatest challenges medical professionals in Africa are faced with is how to incorporate cross-cultural understandings of mental disorders into psychiatric screening of their patients. A majority of psychiatric models of symptomatology and the screening tools developed as a result

of them, are based on Western behavioral norms and thus are not necessarily applicable to other cultures (Akyeampong, Hill, & Kleinman, 2015). Cultural validity, as defined by Solano-Flores & Nelson-Barber (2001), measures the success with which mental disorder screening tools incorporate socio-cultural influences that shape thinking and responses to items. These influences include, but are not restricted to, the pre-existing values, attitudes, encounters, communication patterns, teaching and learning styles, and the philosophical theory of knowledge that are fundamentally in people's cultural backgrounds. These cultural nuances underscore a cultural group's internal attributes but are also fundamental to their perception of external cultures and notions, including mental health (Solano-Flores & Nelson-Barber, 2001).

Communication patterns are play a large role in a groups cultural influences especially pertaining to terms specific to a population. In sub-Saharan Africa, care for individuals continues to be hindered by language barriers. Adapting screening tools into local terminology is a primary and fundamental step for carrying out research on mental health or conducting screening in other cultures, as a lot of concepts are not easily translated across languages. In many cultures, the terms for mental symptoms are different from terms used for the same symptoms in other cultures. This is unlike physical symptoms of common illnesses which are easily transcribed. Thus, to ensure accuracy in diagnosis, it is imperative that when screening tools are used for a population it was not developed for, that the population is able to fully understand the concepts being measured when the tool is administered (Akyeampong, Hill, & Kleinman, 2015). According to World Health Organization (2010), back translation is deemed the best process to create different language versions of an English instrument so that it is conceptually identical for the target population. It involves the process of a person or group of people translating a tool into

a target language followed by a different group then translating it back into English, the original tool and the translated tool are then compared for similarities. This ensures that expressions are equivalent in both languages. The process of back translation is strengthened by the forward translation being conducted by bilingual individuals, preferably those who are familiar with the subject matter covered by the instrument and also with knowledge of the nuances of the target population. Back translation -the process of re-translating the tool into English- should also be undertaken set of bilingual individuals, different from the original translators, whose first language is English and who can also understand the conceptual equivalence of both translations. If conceptual equivalence is not returned in the re-translated English version, then the instrument fails to sufficiently meet the standard of back translation and would therefore need to undergo additional processes until it satisfies this requirement (World Health Organization, 2010).

Lastly, due to Africa's colonial history, there still remains many patriarchal and religious structures in place in African societies. Research highlights that gender dictates the power and control men and women have over different socioeconomic determinants of their mental health, and how they are subsequently treated as a result of this in society ("Gender and women's mental health", 2013). Finally, sub-Saharan African countries face large economic burdens with a poverty rate of 41% and a literacy rate of approximately 65% for people aged 15 and above (The World Bank, n.d.).

The cultural influences mentioned above – language barriers, socioeconomic context of the population, colonial history, religion and gender roles – are some important, but not extensive

factors to consider when discussing the cultural validity of assessment tools in sub-Saharan Africa.

Research Gap

There have been four previous reviews on the validity of screening tools conducted in sub-Saharan Africa. However, three of these reviews focused on the tools used in the assessment of a particular disorder or a specific population as opposed to the full spectrum of mental disorders. They were: depression among HIV-positive persons (Tsai, 2014), perinatal depression (Tsai et al., 2013) and depression among youths (Mutumba, Tomlinson & Tsai, 2014). The fourth review was broader and was conducted by Ali, Ryan & Silva (2016) to produce an extensive summary of common mental disorder screening tools that had been validated in low-and-middle-income countries, but like the other reviews, did not focus on the cultural validity of the tools in their assessments but rather discussed the diagnostic odds ratio as a measure of instrument validity. All of these reviews, although beneficial to the body of literature on mental health in sub-Saharan Africa, don't focus on or incorporate cultural validity in their measure of instrument validity. Therefore, no extensive reviews exist that highlight the cultural validity of screening instruments, for all mental disorders, in all sub-Saharan African countries.

This gap in knowledge of the current attempts at ensuring cultural validity in screening instruments, of all mental disorders, in sub-Saharan Africa, needs to be filled in order to improve the future of mental health care in the region. It is only with culturally valid tools that are able to produce true positives and thus allow for early screening and subsequently early treatment that the mental health burden in sub-Saharan Africa can begin to improve.

II. Background and Literature Review

The World Health Organization (2018) defines mental disorders as one or more "abnormal thoughts, perceptions, emotions, behavior and relationships with others". Mental health services are widely underfunded in countries all over the world but more so in developing countries. Only 28% of countries worldwide have separate provisions in their health budgets for mental health care. Of the developing countries that do have such budgets, 62% of them spend under 1% of their total health budgets on mental health. Thus, there is a significant gap between the burden of mental disorders on societies and the resources dedicated to mental health services (Funk, Saraceno, Pathare, Flisher, & World Health Organization, 2003).

In order to understand the origins of the disparities in care that exist between developed and developing countries, the historical perspective towards diagnosis and treatment must be profiled. Western-type mental health services in sub-Saharan Africa began with colonial powers constructing mental health institutions during the late 19th and early 20th centuries. At the time, mental disorders in sub-Saharan Africa were being diagnosed through evidence of an impaired physical state and those deemed impaired were committed to these institutions (Funk, Saraceno, Pathare, Flisher, & World Health Organization, 2003). However, at the same time, experimental psychology was beginning to develop in the Western world with the opening of Wilhelm Wundt's experimental lab in Germany in 1879. Wundt's institution attempted to measure individual differences in certain measures, in order to profile intelligence. Following these experiments, British psychologist, Sir Francis Galton began to develop tests to study intelligence using Wundt's established measures. Formal psychological tests thus stemmed from these efforts and although they began to be adopted in Western diagnosis. The same was not true for sub-

Saharan Africa where obvious physical impairment was still a key criterion for diagnosis (Stiffler & Dever, 2015). With the rise of formal psychological tests came about a shift from hospital-based care to community-based systems (Funk, Saraceno, Pathare, Flisher, & World Health Organization, 2003). This shift in deinstitutionalization coincided with the World Health Organization framework for effective screening (Wilson, Jungner, & World Health Organization, 1968) which helped to frame the process of mental health screening. These two factors - psychological tests and deinstitutionalization - heralded in the era of regular mental health screening in the West and transformed mental health services to focus on early diagnosis and quick referral of individuals into care (Stiffler & Dever, 2015). This was not the case in sub-Saharan Africa and mental health screening tools did not become frequent until the mid 1990s. Today the frequency and utility of screening tools in diagnosing mental health disorders in sub-Saharan Africa is still fragmented thus why the gap highlighted above still persists (Funk, Saraceno, Pathare, Flisher, & World Health Organization, 2003).

Although screening tools are more frequently used in sub-Saharan Africa today, only a small number of studies have sought to categorize their validity in the African population. Specifically, only four studies have attempted to systematically review studies on screening tools. All of these reviews illustrated that a large proportion of screening tools in use were developed in Western countries so there was variability in instrument positive and negative classifications depending on the target population (Tsai, 2014; Tsai et al., 2013; Mutumba, Tomlinson & Tsai, 2014; Ali, Ryan & Silva, 2016). One went on to note that this variability is due to the lack of standard cut-off scores and that screening tools should undergo pilot tests and be measured against a gold standard (Ali, Ryan & Silva, 2016). Another study raised a similar point and noted the lack of

locally relevant criterion standards but that more research is needed on this before validation studies are undertaken (Mutumba, Tomlinson & Tsai, 2014; Tsai 2013). Despite concluding that the resulting tools were sub-optimal for measuring mental disorders in sub-Saharan Africa, none of the reviews described exact methodologies of how the tools are being administered (Tsai, 2014; Tsai et al., 2013; Mutumba, Tomlinson & Tsai, 2014; Ali, Ryan & Silva, 2016).

There is an emergent need to systematize and strengthen screening tools in sub-Saharan Africa. This is due to the revitalized focus on the role cultural constructs impact a populations' understandings of mental health disorders in order to help decrease the burden of mental disorders. Literature reveals a wealth of studies on mental disorder screening tools. This review methodically analyzes and identifies the cultural constructs that are being integrated into screening tools to make them valid for use in sub-Saharan Africa.

III. Methods

3.1 Study Aim

This paper aims to conduct and report on a high-quality systematic review of studies that assess the validity of screening instruments for mental disorders in sub-Saharan Africa, and to critically appraise the construct and cultural validity utilized. Construct validity defines how well a tool measures a specific theoretical construct and cultural validity, as highlighted above, is a measure of the ways a tool incorporates or accounts for the influences of a particular culture. These outcome measures will seek to understand the reliability, validity, sensitivity and specificity to the community and population they seek to aid.

3.2 Literature Search Strategy

The following electronic databases were chosen with the plan to discover pertinent subject matter in credible papers: 1) Pubmed, 2) Web of Science, 3) Popline, 4) Cochrane Library, 5) PsychINFO. These databases were selected as they are common databases for health sciences research and are regularly used by global, public health experts. The databases contain either high quality peer-reviewed articles published in journals or graduate level dissertations, all with broad linkages to similar high-profile databases and because they incorporate work from the Sub-Saharan African region. Only original research papers were included. Six different searches were conducted using terms, displayed in Table 1 below, to identify relevant papers. Synonymous terms were identified using the MeSH database. Broad terms for mental health were used to ensure a broad scope was captured as studies were theorized to likely mention "mental health" in the background of the paper even if more specific cultural terms are referenced as well. To minimize bias caused by human error, each database had the search

repeated three times to ensure the same number of results were generated each time (Egger, & Smith, 1998).

Table 1: Literature Search Strategy

Search 1	Search 2-6
Mental Health **or** Mental Disorder **or** Mental Illness **and** Title: Test **or** Assess* **or** Instrument **or** Screening **or** Tool **and** Africa **and** Valid* **and** Cultur*	Mental Health **or** Mental Disorder **or** Mental Illness **and** Title: Screening (2) Title: Tool (3) Title: Instrument (4) Title: Assess* (5) Title: Test (6) **and** Africa

3.3 Scope of Review

The following definitions were utilized for the terms that appear in this review henceforth. Definitions were selected from organization who are major global health actors and are reliable sources for public health experts:

1. Mental Health: This review defined Mental health based on the World Health Organization's (2014) definition of a "a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community"

2. Mental Disorders (or Mental Illness): Mental disorders were defined based on the U.S. National Library of Medicine's terminology, which defines it as "conditions that affect your thinking, feeling, mood, and behavior which may be occasional or long-lasting (chronic)" (U.S National Library of Medicine, 2019)
3. Mental Health Screening: Mental Health Screening was defined based on the U.S. National Library of Medicine definition of "An exam of one's emotional health that helps find out if one has a mental disorder" (U.S. National Library of Medicine, 2019)

3.4 Study Selection

All records returned by the search had the abstract reviewed for a possible inclusion match. For those deemed possibly significant, full texts were then retrieved for further evaluation using the inclusion criteria described below. Articles that met the inclusion criteria also had their reference list reviewed to identify additional articles. The author acknowledges the possibility of inclusion bias arising from the criteria for study inclusion.

3.5 Inclusion Criteria

- Study Design: Papers were suitable for selection in this review if they reported the aim of the study as a validation study of one or more mental disorder screening tools
- Disorders: Studies for all mental disorders, except mental disorders due to known physiological conditions, that were listed in the World Health Organization's International Classification of Diseases, version 10, were included (World Health Organization, 2004)
- Location: Studies conducted in Sub-Saharan African countries, according to World Bank's classification, were included (The World Bank, n.d.)

- Population: Populations of all socioeconomic status, education level, gender and age were included
- Language: Studies conducted in any language were included, but publication had to be in English
- Timing: Studies identified that were published by January 1st, 2019 were included

3.6 Exclusion Criteria

So as to maximize the breadth of this review, all studies meeting the above inclusion criteria were included unless they also met the following exceptions: Studies published in non-peer reviewed journals, systematic review articles, studies whose main aim was to develop a screening tool for a specific context or conduct an intervention trial and studies conducted on an African population but in countries outside Sub-Saharan Africa.

IV. Results

4.1 Study Selection

The study selection procedure for this review is illustrated in figure 1 below. Search on the databases produced 2,805 peer-reviewed, published articles and dissertation titles which were reviewed in this study. Only PMC and Web of Science produced results that were eligible for this study. The titles did not warrant further review if they obviously did not evaluate a mental disorder, were not conducted in sub-Saharan African or were studying neurocognitive disorders. Of the 2,805 titles, 373 articles warranted further analysis and thus had their abstracts and study aims reviewed for the inclusion eligibility. 318 articles failed to meet the inclusion eligibility detailed above, leaving 55 studies for inclusion in this review. Two studies were conducted and published in one paper (Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt, 2015) but will be treated as two separate studies for the remainder of this review.

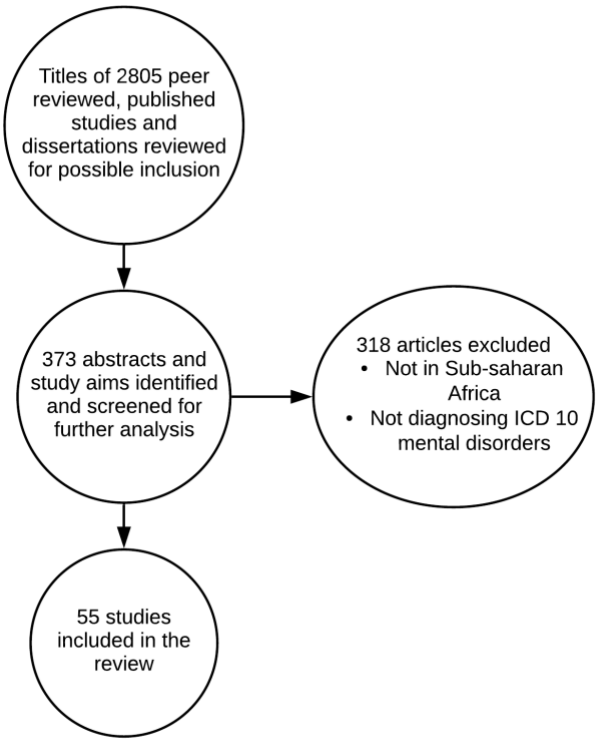

Figure 1: Study Selection Procedure

4.2 Study Features

Most of the included studies (87%; N=48) were published in the past 10-year period (2009-2019). The othere studies were: Adewuya, Ola, & Afolabi (2006); Berard, Boermeester, & Viljoen (1998); Bertschy, Viel, & Ahyi (1992); Muris et al. (2006); Odenwald et al. (2007); Suliman, Kaminer, Seedat, & Stein (2005. The included studies were conducted in 17 different sub-Saharan African countries, with more than a third of studies being from South Africa (38%; N=21) as illustrated in table 2 below.

Table 2: Table of Studies Selected, by Country

Studies	Countries	# of Studies Included from PMC	# of Studies Included from Web of Science
Bertschy, Viel, & Ahyi (1992)	Benin	0	1
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Burundi	0	1
Pence et al. (2012)	Cameroon	0	1
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	Cote d'Ivoire	0	1
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	Ghana	0	1
Hanlon et al. (2015); Gelaye et al. (2014); Geibel et al., (2016); Woldetensay et al. (2018); Habtamu et al. (2017); Gelaye et al. (2013)	Ethiopia	5	1
Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride (2018); Owoso et al. (2014); Denckla et al. (2017)	Kenya	1	2
Udedi, Muula, Stewart, & Pence (2019); Stewart et al. (2009); Stewart, Umar, Tomenson, & Creed (2013); Chorwe-Sungani & Chipps (2018)	Malawi	1	3
Seth et al. (2015)	Namibia	0	1
Marquer et al. (2012)	Niger	1	0
Adewuya, Ola, & Afolabi (2006); Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018); Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	Nigeria	2	1

Studies	Countries	# of Studies Included from PMC	# of Studies Included from Web of Science
Odenwald et al. (2007)	Somaliland	1	0
Shanahan, Anderson, & Mkhize (2001); Berard, Boermeester, & Viljoen (1998); Muris et al. (2006); Sharp et al. (2014); Martin, Fincham, & Kagee (2009); Flisher, Sorsdahl, & Lund (2012); Andersen et al. (2011); Tomita, Kandolo, Susser & Burns (2016); Morojele et al., (2016); Saal, Kagee, & Bantjes (2018); Kagee, Bantjes, Saal, & Sefatsa (2019); Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016); Cholera et al. (2014); Suliman, Kaminer, Seedat, & Stein (2005); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018); Carney, Myers, & Louw (2016); Mellins et al., (2018); Ogle, Koen, & Niehaus (2018); Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015); Mellins et al. (2017)	South Africa	7	13
Nolan et al. (2018); Francis, Helander, Kapiga, Weiss, & Grosskurth (2015)	Tanzania	2	0
Betancourt, Yang, Bolton, & Normand (2014); Akena, Joska, Obuku, & Stein (2013); Blair et al. (2017); Betancourt et al. (2009); Natamba et al. (2014)	Uganda	2	3
Kane, Murray, Bass, Johnson, & Bolton (2016); Chishinga et al. (2011)	Zambia	2	0
Chibanda et al. (2016); January & Chimbari (2018)	Zimbabwe	1	1

The studies evaluated 14 different disorders a total of 108 times, as most studies measured multiple disorders. Mood disorders were the most frequently studied (N= 39), followed by other anxiety disorders (N= 16) and then posttraumatic stress disorder (N= 10). The full list of disorders evaluated in the studies is highlighted in table 3 below. There was a diverse population

measured in terms of age, demographics and gender. Sample sizes ranged from 61 in the smallest study to 4077 in the largest, with an average sample size of 505.

Table 3: Table of Studies Selected, by Disorder and Frequency of Occurrences

Studies	Disorders (ICD 10 Code)	# of Studies
Denckla et al. (2017); Flisher, Sorsdahl, & Lund (2012); Geibel et al., (2016); Sharp et al. (2014)	Attention-deficit Hyperactivity Disorders (F90)	4
Betancourt et al. (2009); Denckla et al. (2017); Flisher, Sorsdahl, & Lund (2012); Geibel et al., (2016); Shanahan, Anderson, & Mkhize (2001); Sharp et al. (2014)	Conduct Disorders (F91)	6
Betancourt, Yang, Bolton, & Normand (2014); Mellins et al. (2018)	Disorders of social functioning with onset specific to childhood and adolescence (F94)	2
Betancourt, Yang, Bolton, & Normand (2014); Marquer et al. (2012); Mellins et al. (2018); Shanahan, Anderson, & Mkhize (2001)	Emotional Disorders with Onset Specific to Childhood (F93)	4
Blair et al. (2017); Carney, Myers, & Louw (2016); Chishinga et al. (2011); Francis, Helander, Kapiga, Weiss, & Grosskurth (2015); Kane, Murray, Bass, Johnson, & Bolton (2016); Mellins et al. (2017); Morojele et al. (2016); Seth et al. (2015); Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016)	Mental and behavioral Disorders due to Psychoactive Substance Use (F10-F19)	9
Habtamu et al. (2017); Owoso et al. (2014); Shanahan, Anderson, & Mkhize (2001)	Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders (F20-F29)	3
Denckla et al. (2017); Geibel et al. (2016); Marquer et al. (2012); Shanahan, Anderson, & Mkhize (2001)	Somatoform Disorders (F45)	4
Habtamu et al. (2017); Owoso et al. (2014); Shanahan, Anderson, & Mkhize (2001)	Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders (F20-F29)	3

Studies	Disorders (ICD 10 Code)	# of Studies
Adewuya, Ola, & Afolabi (2006); Akena, Joska, Obuku, & Stein (2013); Andersen et al. (2011); Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015); Berard, Boermeester, & Viljoen (1998); Bertschy, Viel, & Ahyi (1992); Betancourt et al. (2009); Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015); Blair et al. (2017); Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017); Chibanda et al. (2016); Chishinga et al. (2011); Cholera et al. (2014); Chorwe-Sungani & Chipps (2018); Denckla et al. (2017); Flisher, Sorsdahl, & Lund (2012); Geibel et al., (2016); Gelaye et al. (2013); Gelaye et al. (2014); Habtamu et al. (2017); Hanlon et al. (2015); January & Chimbari (2018); Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride (2018); Marquer et al. (2012); Mellins et al. (2017); Natamba et al. (2014); Nolan et al. (2018); Ogle, Koen, & Niehaus (2018); Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017); Pence et al. (2012); Saal, Kagee, & Bantjes (2018); Shanahan, Anderson, & Mkhize (2001); Sharp et al. (2014); Stewart et al. (2009); Stewart, Umar, Tomenson, & Creed (2013); Tomita, Kandolo, Susser, & Burns (2016); Udedi, Muula, Stewart, & Pence (2019); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018); Woldetensay et al. (2018)	Mood Disorders (F30-F39)	39
Andersen et al. (2011); Betancourt et al. (2009); Chibanda et al. (2016); Denckla et al. (2017); Flisher, Sorsdahl, & Lund (2012); Geibel et al., (2016); Marquer et al. (2012); Mellins et al. (2017); Muris et al. (2006); Ogle, Koen, & Niehaus (2018); Shanahan, Anderson, & Mkhize (2001); Sharp et al. (2014); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Other Anxiety Disorders (F41)	13
Blair et al. (2017); Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017); Denckla et al. (2017); Kagee, Bantjes, Saal, & Sefatsa (2019); Marquer et al. (2012); Martin, Fincham, & Kagee (2009); Mellins et al. (2017); Odenwald et al. (2007); Ogle, Koen, & Niehaus (2018); Suliman, Kaminer, Seedat, & Stein (2005)	Post-traumatic Stress Disorder (F43.1)	10
Flisher, Sorsdahl, & Lund (2012); Marquer et al. (2012); Ogle, Koen, & Niehaus (2018); Shanahan, Anderson, & Mkhize (2001)	Phobic Anxiety Disorders (F40)	4

Studies	Disorders (ICD 10 Code)	# of Studies
Denckla et al. (2017); Shanahan, Anderson, & Mkhize (2001)	Obsessive-compulsive Disorder (F42)	2

4.3 Screening Tools

The studies validated 51 different screening tools which are highlighted in Table 4, by the name of screening instrument and number of studies it was utilized in.

Table 4: Table of Studies Selected, by Type of Screening Tools and Frequency of Occurrences

Studies	Screening Tools	# of Studies
Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	26-items Health Related Quality of Life in Stroke Patients (HRQOLISP-26)	1
Chorwe-Sungani & Chipps (2018)	3-Item screener for depression	1
Habtamu et al. (2017)	36-Item WHODAS 2.0	1
Betancourt et al. (2009)	Acholi Psychosocial Assessment Instrument (APAI)	1
Betancourt, Yang, Bolton, & Normand (2014)	African Youth Psychosocial Assessment (AYPA)	1
Blair et al. (2017); Chishinga et al. (2011); Francis, Helander, Kapiga, Weiss, & Grosskurth (2015); Morojele et al. (2016); Seth et al. (2015)	Alcohol Use Disorders Identification Test (AUDIT)	5
Kane, Murray, Bass, Johnson, & Bolton (2016); Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016)	Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST)	2
Berard, Boermeester, & Viljoen (1998); Saal, Kagee, & Bantjes (2018)	Beck Depression Inventory (BDI)	3
Akena, Joska, Obuku, & Stein (2013); Chishinga et al. (2011); January & Chimbari (2018); Natamba et al. (2014)	Center for Epidemiologic Studies Depression (CES-D) scale	4
Suliman, Kaminer, Seedat, & Stein (2005)	Child and Adolescent Trauma Survey (CATS)	1

Studies	Screening Tools	# of Studies
Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride (2018)	Child Depression Inventory (CDI)	1
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Child PTSD Symptom Scale (CPSS)	1
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Childhood Trauma Questionnaire (CTQ)	1
Mellins et al. (2017)	Client Diagnostic Questionnaire (CDQ)	1
Bertschy, Viel, & Ahyi (1992)	Comprehensive Psychopathological Rating Scale (CPRS)	1
Carney, Myers, & Louw (2016)	CRAFFT Screening Interview	1
Flisher, Sorsdahl, & Lund (2012)	Diagnostic Interview Schedule for Children (DISC-IV)	1
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Depression Self-Rating Scale (DSRS)	1
Chorwe-Sungani & Chipps (2018); January & Chimbari (2018); Stewart, Umar, Tomenson, & Creed (2013); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Edinburgh Postnatal Depression Scale (EPDS)	4
Muris et al. (2006)	Egna Minnen Beträffande Uppfostran for Children	1
Chibanda et al. (2016)	Generalized Anxiety Disorder questionnaire (GAD-7)	1
Carney, Myers, & Louw (2016)	Global Appraisal of Individual Needs-Short Screener (GAIN-SS) substance use subscale	1
Blair et al. (2017)	Harvard Trauma Questionnaire (HTQ)	1
Blair et al. (2017)	Hopkins Symptoms Checklist-15 (HSCL-15)	1
Berard, Boermeester, & Viljoen (1998)	Hospital Anxiety and Depression Scale (HADS)	1

Studies	Screening Tools	# of Studies
Chorwe-Sungani & Chipps (2018)	Hopkins Symptoms Checklist-25 (HSCL-25)	1
Akena, Joska, Obuku, & Stein (2013); Andersen et al. (2011); Hanlon et al. (2015); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Kessler Psychological Distress scale (K10)	4
Andersen et al. (2011); Hanlon et al. (2015); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Kessler Psychological Distress scale (K6)	3
Denckla et al. (2017)	Ndetei-Othieno-Kathuku Scale	1
Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018)	Obsessive-Compulsive Inventory-Revised (OCI-R)	1
Chibanda et al. (2016); Hanlon et al. (2015); Nolan et al. (2018)	Patient Health Questionnaire-2 (PHQ-2)	3
Adewuya, Ola, & Afolabi (2006); Akena, Joska, Obuku, & Stein (2013); Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015); Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015); Chibanda et al. (2016); Cholera et al. (2014); Gelaye et al. (2013); Gelaye et al. (2014); Hanlon et al. (2015); January & Chimbari (2018); Pence et al. (2012); Udedi, Muula, Stewart, & Pence (2019); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018); Woldetensay et al. (2018)	Patient Health Questionnaire-9 (PHQ-9)	14
Carney, Myers, & Louw (2016)	Personal Experience Screening Questionnaire (PESQ)	1
Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018)	Positive and Negative Syndrome Scale (PANSS)	1
Martin, Fincham, & Kagee (2009); Odenwald et al. (2007)	Posttraumatic Diagnostic Scale (PDS)	2
Kagee, Bantjes, Saal, & Sefatsa (2019)	Posttraumatic Symptom Scale (PSS)	1
Chorwe-Sungani & Chipps (2018)	Pregnancy Risk Questionnaire (PRQ)	1
Owoso et al. (2014)	PRIME Screen	1

Studies	Screening Tools	# of Studies
Marquer et al. (2012)	Psychological Screening for Young Children aged 3 to 6 (PSYCAa3-6)	1
Tomita, Kandolo, Susser, & Burns (2016)	Quick Inventory of Depressive Symptomatology (QIDS)	1
Muris et al. (2006)	Screen for Child Anxiety Related Disorders (SCARED)	1
Chorwe-Sungani & Chipps (2018); Hanlon et al. (2015); Stewart et al. (2009); Stewart, Umar, Tomenson, & Creed (2013)	Self-Reporting Questionnaire (SRQ)	4
Chibanda et al. (2016)	Shona Symptom Questionnaire for common mental disorders (SSQ-14)	1
Mellins et al. (2018); Sharp et al. (2014)	Strengths and Difficulties Questionnaire (SDQ)	2
Shanahan, Anderson, & Mkhize (2001)	Symptom Checklist-90-Revised (SCL-90-R)	2
van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Two-item Generalised Anxiety Disorder scale (GAD-2)	1
Ogle, Koen, & Niehaus (2018)	Visual Screening Tool for Anxiety Disorders and Depression (VISTAD)	1
Nolan et al. (2018)	WHO Well Being Index (WHO-5)	1
van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Whooley questions	1
Geibel et al., (2016)	Youth Self-Report (YSR)	1

Aside from the primary instruments highlighted above, 41 of the 55 studies also made use of a reference instrument. The most common reference instruments were the Mini International Neuropsychiatric Interview (26%; N=11) and the Diagnostic interview (the Structured Clinical

Interview for DSM-IV) (SCID) (12%; N=7). A full list of the reference instruments is shown in table 5 below.

Table 5: Table of Studies Selected, by Reference Screening Tool and Frequency of Occurrences

Studies	Reference Standard	# of Studies
Morojele et al. (2016)	AUDIT-10	2
Habtamu et al. (2017)	Brief Psychiatric Rating Scale (BPRS-E)	1
Mellins et al. (2017)	Client Diagnostic Questionnaire (CDQ) clinical version	1
Gelaye et al. (2013); Gelaye et al. (2014)	Clinical Assessment in Neuropsychiatry (SCAN)	2
Berard, Boermeester, & Viljoen (1998); Marquer et al., (2012); Tomita, Kandolo, Susser, & Burns (2016)	Clinical Interview	3
Andersen et al. (2011); Gelaye et al. (2014); Martin, Fincham, & Kagee (2009); Odenwald et al. (2007); Pence et al. (2012)	Composite International Diagnostic Interview (CIDI)	5
January & Chimbari (2018)	Diagnostic and Statistical Manual for Mental Disorders (DSM-5)	1
Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015); Chibanda et al. (2016); Kagee, Bantjes, Saal, & Sefatsa (2019); Saal, Kagee, & Bantjes (2018); Stewart et al. (2009); Stewart, Umar, Tomenson, & Creed (2013); Udedi, Muula, Stewart, & Pence (2019)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	7
Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	Hospital Anxiety and Depression Scale (HADS-D)	1
Suliman, Kaminer, Seedat, & Stein (2005)	Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL)	1
Adewuya, Ola, & Afolabi (2006); Akena, Joska, Obuku, & Stein (2013); Chishinga et al. (2011); Cholera et al. (2014); Francis, Helander, Kapiga, Weiss, & Grosskurth (2015); Hanlon et al. (2015); Natamba et al. (2014); Ogle, Koen, & Niehaus (2018); Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018); van Heyningen, Honikman, Tomlinson, Field, & Myer (2018); Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016); Woldetensay et al. (2018)	Mini International Neuropsychiatric Interview (MINI)	11
Sharp et al. (2014)	NIMH Diagnostic Interview Schedule for Children (CDISC-IV)	1
Nolan et al. (2018)	Patient Health Questionnaire-9	1
Owoso et al. (2014)	Psychosis-Risk Syndromes (SIPS)	1
Odenwald et al. (2007)	Self-Report Questionnaire (SRQ-20)	1

Denckla et al. (2017); Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Youth Self Report DSM-IV oriented scales	2

4.4 Cultural Validity

This review identified several techniques that the researchers undertook to ensure the cultural characteristics of the populations were incorporated in their validation work.

4.4.1 Technique 1: Language Translation

25 studies (45%) conducted language translation of the instruments with back translation, into the local language of the study population. The studies, the languages they were translated into and method of translation is summarized in Table 6 below. 2 studies were conducted in a local language with tools that were originally developed in a local language (Betancourt et al., 2009; Betancourt, Yang, Bolton, & Normand, 2014). 8 other studies back translated instruments without mentioning who conducted translations (Cholera et al., 2014; Flisher, Sorsdahl, & Lund, 2012; Gelaye et al., 2014; Kane, Murray, Bass, Johnson, & Bolton, 2016; Mellins et al., 2018; Muris et al., 2006; Nolan et al., 2018; Seth et al., 2015). 6 studies validated tools in the local language but did not undergo the translation process themselves as the tools had been previously translated in other studies (Bertschy, Viel, & Ahyi, 1992; Chorwe-Sungani & Chipps, 2018; Habtamu et al., 2017; Hanlon et al., 2015; January & Chimbari, 2018; Stewart, Umar, Tomenson, & Creed, 2013). 11 studies conducted screenings in English, as it was deemed the common language of communication for that population (Adewuya, Ola, & Afolabi, 2006; Kagee, Bantjes, Saal, & Sefatsa, 2019; Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride, 2018; Martin, Fincham, & Kagee, 2009; Mellins et al., 2018; Muris et al., 2006; Owoso et al., 2014; Pence et al., 2012; Saal, Kagee, & Bantjes, 2018; Seth et al., 2015; Tomita, Kandolo, Susser, & Burns, 2016). 1 of these 11 studies had a bilingual project staff member present during

screening for translation if needed (Carney, Myers, & Louw; 2016). Lastly, 9 studies did not mention at all if any translation occurred or was needed (Akena, Joska, Obuku, & Stein, 2013; Berard, Boermeester, & Viljoen, 1998; Francis, Helander, Kapiga, Weiss, & Grosskurth, 2015; Morojele et al., 2016; Ogle, Koen, & Niehaus, 2018; Ojagbemi, Owolabi, Akinyemi, & Ovbiagele, 2017; Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye, 2018; Suliman, Kaminer, Seedat, & Stein, 2005; Westhuizen, Wyatt, Williams, Stein, & Sorsdahl, 2016).

Table 6: Table of Studies that Conducted Translations, Language of Translation and Technique

Studies	Language Translation	Translation Technique
Andersen et al. (2011)	Yes, to Afrikaans, Zulu, Xhosa, Northern Sotho, and Tswana	By bilingual and multilingual experts
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	Yes, to French (CDI) and Twi (Ghana)	By health care professionals
Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015)	Yes, to seTswana	By a seTswana-speaking mental health professional; back-translated by an independent seTswana-speaking clinical psychologist
Blair et al. (2017)	Yes, to Acholi Luo	By experienced Acholi researchers
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Yes, from French to Kirundi	By the interviewers, who were bilingual in both the languages
Chibanda et al. (2016)	Yes, the PHQ-9 and GAD, to Shona	By a different independent language expert.
Chishinga et al. (2011)	Yes, to Bemba, Nyanja, Tonga and Lozi	By professional translators working for the Zambia National Broadcasting Corporation (ZNBC)
Chorwe-Sungani & Chipps (2018)	Yes, only the HSCL-15, PRQ and 3-item screener for depression.	By the first author and a social worker
Denckla et al. (2017)	Yes, to Kiswahili and Kikamba	By an experienced health professional, back translation by an independent team blind to the original instrument
Geibel et al., (2016)	Yes, to Amharic	By a psychologist; back-translated to English by a professional translator
Gelaye et al. (2013)	Yes, to Amharic	By the lead author; back translation by panels of bilingual experts
Hanlon et al. (2015)	Yes, PHQ-9, to Amharic	By two Ethiopian psychiatrists
Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride	Yes, to Kiswahili	By a team of experts drawn from the study team

Studies	Language Translation	Translation Technique
(2018)		
Marquer et al. (2012)	Yes, to Hausa	By two professional translators, independently
Mellins et al. (2017)	Yes, to isiZulu	By a bilingual isiZulu-English South African lay counselor, back-translation by the SA psychologist
Natamba et al. (2014)	Yes, to Acholi and Langi	By local research staff
Odenwald et al. (2007)	Yes, to Somali	By independent professional translators
Shanahan, Anderson, & Mkhize (2001)	Yes, to Zulu	By Zulu psychologists
Sharp et al. (2014)	Yes, to Sesotho	By two native Sesotho speakers, and a third native speaker then back-translated
Stewart et al. (2009)	Yes, to Chichewa	By 4 bilingual Malawians: a psychiatrist (FK), a clinical psychologist (EU), a psychiatric nurse and a secondary school teacher.
Stewart, Umar, Tomenson, & Creed (2013)	Yes, to Chichewa; SRQ and SCID previously translated	By a Malawian clinical psychologist, back-translated by a bilingual independent non-mental health professional.
Udedi, Muula, Stewart, & Pence (2019)	Yes, to Chichewa	By two bilingual Malawians: one mental health nurse and one linguistics and communication specialist; By independent bilingual Malawians back-translated
van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Yes, to isiXhosa and Afrikaans	By translators at Stellenbosch University Language Centre
Woldetensay et al. (2018)	Yes, to Wellega and Tulama Afaan Oromo	By two native speaker mental health specialists

4.4.2 Technique 2: Terms Validation and Adaptation

Another cultural validation technique employed was adaptation of specific terms in the tools, and this was conducted using three different techniques: group consensus meeting, focus groups and pre-study pilot testing. 28 studies (51%), of the 55, conducted these terms validation exercises which are specifically detailed in table 7 below. Of these, 13 studies (46%) conducted group meetings with the translators, research team and/or external experts to come to a group consensus on any items that posed translation issues during the process of back translations (Andersen et al., 2011; Bertschy, Viel, & Ahyi, 1992; Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017; Chibanda et al., 2016; Mellins et al., 2017; Hanlon et al., 2015; Natamba et al.,

2014; Odenwald et al., 2007; Ojagbemi, Owolabi, Akinyemi, & Ovbiagele, 2017; Shanahan, Anderson, & Mkhize, 2001; Sharp et al., 2014; Woldetensay et al., 2018; Owoso et al., 2014), 3 studies (11%) conducted a focus group with a small sample of the population to understand their perceptions of mental health and thus to inform alterations of the language of the tools by incorporating specific descriptions and mental health terms known to that population (Betancourt et al., 2009; Kane, Murray, Bass, Johnson, & Bolton, 2016; Pence et al., 2012) and 3 studies (11%) pilot tested the translated tools in a small sub sample of the population and further reconstructed language and other constructs based on the pilot results (Blair et al., 2017; Kagee, Bantjes, Saal, & Sefatsa, 2019; Saal, Kagee, & Bantjes, 2018). Finally of the remaining 9 studies: 4 studies conducted both a focus group and had a consensus meeting (Habtamu et al., 2017; Mellins et al., 2018; Stewart et al., 2009; Udedi, Muula, Stewart, & Pence, 2019), 3 studies pilot tested the translated tool and had a consensus meeting (Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt, 2015; Chishinga et al., 2011; Denckla et al., 2017) and 2 studies conducted the focus groups, a pilot test as well as consensus meeting (Marquer et al., 2012; Stewart, Umar, Tomenson, & Creed, 2013). One other study made changes based on previous research (Cholera et al., 2014).

Table 7: Table of Studies that Conducted Translations, Language of Translation and Technique

Study	Terms Validation Methods	Terms Validation Results
Andersen et al. (2011)	Discrepancies that emerged from the back-translation were decided on by consensus of a team of experts	Corrections were made as needed
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	Both language versions were reviewed by a committee to ensure their equivalence; Questionnaires were pilot-tested to ensure the comprehensibility of the items.	Corrections were made as needed
Bertschy, Viel, & Ahyi (1992)	Proposed a modified version of the DSM III criteria for major depression; During the four first months the author explored the feasibility of the clinical rating with the CPRS to become familiar with the way in which Beninese people express what they feel, to evaluate the local meaning of certain words or the understanding of certain concepts; Authors went	Preserved the global structure of the DSM III criteria; Taking into account specificity of depression in Black Africa: depression without dysphoric mood, ideas of persecution and somatic complaints; At the end of this

Study	Terms Validation Methods	Terms Validation Results
	progressively deeper into interviews with a kind of cross- cultural back translation: every time a patient seemed to report a symptom, spontaneously or after having been asked, we asked him to give us a concrete description using several examples	prestudy stage, authors started to make complete ratings with the CPRS
Betancourt et al. (2009)	45 free listing interviews with local youth and adults and 57 in-depth interviews with local key informants to identify and describe the important problems of youth in these camps for internally displaced people (IDPs)	Qualitative study identified eight items indicative of prosocial behavior among youth in this setting
Blair et al. (2017)	Translated by experienced Acholi researchers and tested in a pilot community	-
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Each item in Kirundi was followed by extensive discussions in the group until all members reached an agreement that the items would have semantic, conceptual and content equivalence across different socio-economic strata of the society; Like the CTQ, all the depression and PTSD items were discussed by the group of bilingual professionals, and slightly adapted when the group thought it to be necessary.	The team, after extensive deliberation, added the word 'kenshi' (referring to 'often' in English) to item 24 comprising the subscale of sexual abuse
Chibanda et al. (2016)	Draft Shona version was reviewed by a team focused on ensuring contextual equivalence to the original versions based on their understanding and use of local terms for mental distress including kufungisisa (thinking too much), kusuwa (to feel sad), kusuwisisa (to feel very sad) and mwoyo unorwadza (painful heart); A committee of the first author with the clinical psychologist and counselor examined both the original and back translated versions and resolved any discrepancies by consensus	Corrections were made as needed
Chishinga et al. (2011)	All translated versions of the questionnaires were discussed by the research team comprising of members who were fluent in the native languages; Instrument was tested before use in the field	Discrepancies in conceptual and semantic equivalence was resolved through informal committee consensus
Cholera et al. (2014)	-	Questions about substance abuse and knowledge of HIV status (from prior testing experiences) were also included in the questionnaire.
Denckla et al. (2017)	Discrepancies in translation were reconciled by an expert committee of at least six multilingual experts in English, Kiswahili, and Kikamba; Translated scale was administered to a pilot sample for accuracy	Panel members suggested alternatives and agreed on acceptable solutions

Study	Terms Validation Methods	Terms Validation Results
Habtamu et al. (2017)	Administered the scale to a sample of people with SMD and their caregivers (participants were from rural areas and had primarily level educational) for cognitive interviewing to identify any difficulty with understanding of individual items and response categories, as well as acceptability and burden of the whole scale; Expert panel composed of psychiatrists, psychologists, social workers and mental health researchers suggested how each problematic item should be rephrased to be more easily understood by the respondents, while retaining semantic equivalence with the original scale.	Iterative adjustments were made to the Amharic translations, including the addition of examples to items asking about broad and abstract experiences, and replacing less relevant and uncommon concepts to the setting with equivalent but local experiences; For the item "sexual activities" the Amharic translation was changed to ask about romantic relationships.
Hanlon et al. (2015)	The final version was obtained by expert consensus	-
Kagee, Bantjes, Saal, & Sefatsa (2019)	Consulted with one of the SCID developers, Dr Michael First; Conducted a pilot study to ensure that participants found the SCID items comprehensible and feasible	SCID questions were adjusted so that they corresponded to the DSM5 diagnostic criteria; Adapted the language of the instrument accordingly so that it conformed to the criteria for PTSD as delineated in the latest version of the DSM
Kane, Murray, Bass, Johnson, & Bolton (2016)	All questions were then reviewed by community groups of adolescents from our target population for conceptual understanding and translation accuracy	Corrections were made as needed
Marquer et al. (2012)	The translated tool was administered to a small group of caregivers in a pilot phase for acceptance, adequacy and applicability; Focus group discussions with community key informants, individual interviews and expert meetings were conducted to elucidate understanding of the beliefs and perceptions about child development and expectations of the outcome of treatment; Consultations with psychiatrists and anthropologists and the health care staff of hospitals to define developmental milestones and child psychopathology were also organized.	A list containing misunderstood items developed and an alternate proposed translation was made, the final version was fixed by the two translator's consensus taking into account differences between spoken and written Hausa
Mellins et al. (2017)	US and SA Principal Investigators (PI) reviewed discrepancies for meaning and correct diagnostic translation	Corrections were made as needed
Mellins et al. (2018)	Discussions to resolve discrepancies and reviewed with the SDQ author; Focus groups were conducted with SA native isiZulu speakers;	Corrections were made as needed
Natamba et al. (2014)	Conceptual and semantic equivalence were resolved through discussion involving all the translators, the research assistants, and the GRRH psychiatrist	Corrections were made as needed
Odenwald et al. (2007)	The translation and cultural adaptation were carried out by groups of local bilingual and international experts, who had received education in trauma-related concepts; Group discussions and consultations with external specialists were dedicated to culturally	Adapted to the Somali language, culture, and Islamic religion (Somali-PDS); Repeated back translations as needed

Study	Terms Validation Methods	Terms Validation Results
	specific meanings of items and typical experiences in Somalia	
Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	Consultant psychiatrist independently examined each of the Psycho-emotional items and determined whether the scale contained the minimum number of items required for the diagnostic assessment of depression specified in the ICD 10.	-
Pence et al. (2012)	Four focus groups were conducted with hospital patients and family members, groups were separated by gender and religion (Christian and non-Christian) and were Audio-recorded and transcribed for content analysis; Optional clarifications (that were read only if the original question was not understood)	Analysis suggested minor wording changes to two of the questions and optional clarifications (that were read only if the original question was not understood) for six additional questions.
Saal, Kagee, & Bantjes (2018)	Prior to the commencement of the study, we conducted a pilot study to determine the comprehensibility and feasibility of the SCID among the population of interest;	The results of the pilot study showed that the SCID was suitable for use among individuals seeking HIV testing in South Africa
Shanahan, Anderson, & Mkhize (2001)	The translators were instructed to rate each item according to the difficulty encountered in achieving what they regarded as an equivalent translation; All 90 items, as well as the instructions, response categories, and example, were reviewed; Particular attention was paid to items which had (a) been rated by the translators as presenting problems in achieving equivalent translation (ratings of 2 or 3), and/or (b) resulted in back-translations which were nonequivalent to the original English items; A committee, consisting of the three Zulu psychologists and the psychiatric nurse, met to consider two back-translated Zulu versions of the SCL-90-R and each item was considered individually in the committee, the pros and cons of different Zulu versions of the items being openly debated over several meetings, totalling more than eight hours of committee time	Six items (numbers 30,47,57,64,65, and 86) were identified by the committee as requiring decentering in order to facilitate translation into Zulu; Based on committee discussions, Five SCL-90-R items (numbers 5, 9, 12,62,78) were identified as possibly requiring revision in order to make the items more comprehensible or less offensive to respondents and were revised to the cultural context; For the final tool, 48 items from the first translation were selected, 35 from the second translation and 8 items were rewritten to include a combination of elements from both translations
Sharp et al. (2014)	Native Sesotho translators met following individual translations; All translators met to agree on the final version	-
Stewart et al. (2009)	The 4 translations were reviewed together so variations between the translations of each item were discussed; 3 focus groups were held, each with 4 to 6 participants who were invited to discuss and comment upon the SRQ translation; Pilot SRQ was administered to 12 mothers/guardians of infants admitted to a nutritional rehabilitation unit and problematic items were discussed with the women in order to check how they had understood the item	Consensus reached based on meeting; Further modifications were made on the basis of the back-translations; After each focus group further modification was made to the translation based on the comments of the participants; The interviewers noted any items about which the pilot subjects

Study	Terms Validation Methods	Terms Validation Results
		asked for clarification; Experience from the piloting was discussed by the group and final version was agreed; A number of items presented difficulties in translation, as some of the English terms did not have an exact equivalent in Chichewa and in these cases either the most closely matching word was used, or two words that between them fully captured the English meaning were included, each item was prefaced with "In the last 4 weeks…." and in one case the order of items was changed to avoid misinterpretation
Stewart, Umar, Tomenson, & Creed (2013)	RS advised on the concepts captured by the original English wording of each item to guide the choice of Chichewa expression; Discussion was held with nurses from the antenatal clinic who were invited to comment upon the EPDS translation; Piloted to a number of women attending the antenatal clinic	Translation followed the method used in the development of the SRQ in Chichewa previously described; The choice of Chichewa expression was based on RS advising; Modifications were made on the basis of the back-translation; Further modification was made to the translation based on nurses' comments; Problematic items were discussed with focus group women in order to check understanding and final version was agreed; The use of a visual prompt card occured to facilitate the responses.
Udedi, Muula, Stewart, & Pence (2019)	Evaluation of the translated tools by the principal investigator, two mental health professionals and two health promotion officers with extensive expertise in developing health communication tools; Pretested the Chichewa translated version of the PHQ 9 on a convenience small sample of 15 participants where participants were probed about their perceived interpretation of the constructs	Consensus translation arrived at after evaluation; Pretesting helped to identify any challenges respondents might have with the translation and unclear Chichewa terms were modified to include terms that are more commonly used and understandable by the participants
Woldetensay et al. (2018)	A reconciliation meeting was conducted to identify any areas presenting linguistic problems and to assess the participants' level of understanding	All items were revised for comprehension (meaning and question objectives), information retrieval (type of information and recall strategy), decision process (sensitivity and social desirability), and adequacy of response options)

4.4.1 Technique 3: Method of Administration

The studies also employed various methods of administering the tools. 40 studies (72%) conducted an interview method of administration. Of these, 5 studies utilized local, lay interviewers who were trained in how to properly administer the tool prior to the study (Hanlon et al., 2015; Odenwald et al., 2007; Mellins et al., 2018; Mellins et al., 2017; Betancourt, Yang, Bolton, & Normand, 2014;) and 1 study engaged local, lay interviewers who did not undergo training prior to the study (Habtamu et al., 2017). 4 studies engaged trained health workers such as nurses and psychiatrists (Akena, Joska, Obuku, & Stein, 2013; Bertschy, Viel, & Ahyi, 1992; Geibel et al., 2016; Sharp et al., 2014) and another 4 studies also utilized health workers, but they were not trained before conducting the interviews. Additionally, 13 studies enlisted members of the research team to do the interviewers (Chorwe-Sungani & Chipps; 2018; Flisher, Sorsdahl, & Lund, 2012; Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbridem 2018; Natamba et al., 2014; Nolan et al., 2018) with 8 of these 13 studies conducting interview training techniques prior to conducting the studies (Berard, Boermeester, & Viljoen, 1998; Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Chibanda et al., 2016; Chishinga et al., 2011; Cholera et al., 2014; Udedi, Muula, Stewart, & Pence, 2019; Shanahan, Anderson, & Mkhize, 2001; Pence et al., 2012). 5 studies procured psychology bachelor or master's degree holders or persons with experience in conducting research and had them trained to conduct interviews using the tools in their studies (Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt, 2015; Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017; Westhuizen, Wyatt, Williams, Stein, & Sorsdahl, 2016; van Tomita, Kandolo, Susser, & Burns, 2016; Saal, Kagee, & Bantjes, 2018) and 1 additional study also used these interviewers but without administering any training beforehand (Heyningen, Honikman, Tomlinson, Field, & Myer, 2018). Finally, 11 studies did not mention

who administered the interviews (Martin, Fincham, & Kagee, 2009; Seth et al., 2015; Suliman, Kaminer, Seedat, & Stein, 2005; Woldetensay et al., 2018; Stewart, Umar, Tomenson, & Creed, 2013), but 5 of the 11 did mention the interviewers were trained (Betancourt et al., 2009; Kagee, Bantjes, Saal, & Sefatsa, 2019; Marquer et al., 2012; Morojele et al., 2016). Furthermore 2 studies noted that the instruments were self-administered but that assistance was available from a researcher during completion of the tool if the participants needed it (Kane, Murray, Bass, Johnson, & Bolton, 2016; Carney, Myers, & Louw, 2016). The remaining 13 studies did not mention the method by which the tools were administered in their studies (Adewuya, Ola, & Afolabi, 2006; Andersen et al., 2011; Blair et al., 2017; Denckla et al., 2017; Francis, Helander, Kapiga, Weiss, & Grosskurth, 2015; Gelaye et al., 2013; Gelaye et al., 2014; January & Chimbari, 2018; Muris et al., 2006; Ogle, Koen, & Niehaus, 2018; Ojagbemi, Owolabi, Akinyemi, & Ovbiagele, 2017; Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye, 2018; Stewart et al., 2009). Lastly, only 5 studies noted the method of administration of the reference screening tools and all of them indicated the use of trained psychiatrists to administer the reference tools (Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Chibanda et al., 2016; Hanlon et al., 2015; Owoso et al., 2014; Pence et al., 2012).

4.5 Statistical Analysis

Data analysis was the primary basis for concluding the validity of the instruments being measured in all the studies. As highlighted above, sensitivity and specificity measures can be indicative of instrument external reliability and ability to reproduce the same results correctly. 33 out of 55 studies (60%) developed a Receiver Operating Curve to calculate sensitivity and specificity. The cut-off, sensitivity and specificity values are displayed in table 7 below. The PHQ-9 tool, which was the most frequently studied tool, had a cut-off value of 9 in 3 studies

(Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Adewuya, Ola, & Afolabi, 2006; Udedi, Muula, Stewart, & Pence, 2019) and a cut-off value of 10 in 3 studies (Gelaye et al., 2013; Akena, Joska, Obuku, & Stein, 2013; Cholera et al., 2014; Pence et al., 2012). One study had a cut-off of 8 (Woldetensay et al., 2018) and a another one had a cut-off of 5 (Hanlon et al., 2015). Several studies conducted other statistical tests which are detailed fully in Appendix B.

Table 8 Table of Sensitivity, Specificty and Area under Curve Cutoff Values

Study	Screening Tool	Area Under Curve Cutoff	Sensitivity	Specificity
Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015)	PHQ-9	≥ 9	49%	94%
Chishinga et al. (2011)	CES-D; AUDIT	22 (CES-D); 24 (AUDIT)	73% (CES-D); 60% (AUDIT)	-
Francis, Helander, Kapiga, Weiss, & Grosskurth (2015)	MINI	≥5	12%	94%.
Geibel et al., (2016)	YSR	≥ 6.5	75%	63.1%
Gelaye et al. (2013)	PHQ-9	≥10	71.1%	76.6%
January & Chimbari (2018)	EPDS; PHQ-9; CES-D	-	-	-
Kane, Murray, Bass, Johnson, & Bolton (2016)	ASSIST	1 (Low vs. Moderate); 24 (Moderate vs. High)	67.9 (Low vs. Moderate); 58.5 (Moderate vs. High)	62.7 (Low vs. Moderate); 59.4 (Moderate vs. High)
Mellins et al. (2017)	CDQ	-	73%	81%
Natamba et al. (2014)	CES-D	16; 17	79.6%; 78.5%	75.9%; 72.7%.
Odenwald et al. (2007)	PDS	13/14	90%	79%
Ogle, Koen, & Niehaus (2018)	VISTAD	7	90.91%	60.00%
Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018)	OCI-R	10	90%	66.2%
Woldetensay et al. (2018)	PHQ-9	8	80.8%	79.5%
Adewuya, Ola, & Afolabi (2006)	PHQ-9	5 (Major+Minor Depression); 9 (Major Depression)	89.7% (Major+Minor Depression); 92.3%	98.9% (Major+Minor Depression); 98.4%

Study	Screening Tool	Area Under Curve Cutoff	Sensitivity	Specificity
			(Major Depression)	(Major Depression)
Akena, Joska, Obuku, & Stein (2013)	PHQ-9	10	91%	81%
Andersen et al. (2011)	K6; K10	≥ 10 (K6); ≥ 16 (K10)	70% (K10); 70% (K6)	67% (K10); 62% (K6)
Berard, Boermeester, & Viljoen (1998)	BDI; HADS	16; 8	85% (BD); 71% (HADS)	95% (BDI); 95% (HADS)
Chibanda et al. (2016)	SSQ-14; PHQ-2; PHQ-9; GAD-17	9 (SSQ-14); 11 (PHQ-2); 2 (PHQ-9); 10 (GAD-17)	84% (SSQ-14, any CMD); 85% (PHQ-9, Depression); 91% (PHQ-2, Depression); 89% (GAD-7, Anxiety)	73% (SSQ-14, any CMD); 69% (PHQ-9, Depression); 40% (PHQ-2, Depression); 73% (GAD-7, Anxiety)
Cholera et al. (2014)	PHQ-9	10	78.7%	83.4%
Hanlon et al. (2015)	PHQ-2; PHQ-9; SRQ-20; K6; K10	1 (PHQ-2); 5 (PHQ-9); 8 (SRQ-20); 9 (K6); 18 (K10)	83.3% (PHQ-2); 83.3% (PHQ-9); 83.3% (SRQ-20); 77.8% (K6); 77.8 (K10)	60.8% (PHQ-2); 74.7 (PHQ-9); 74.3% (SRQ-20); 73.3% (K6); 76.7% (K10)
Kagee, Bantjes, Saal, & Sefatsa (2019)	PSS	22.5	76%	78%
Martin, Fincham, & Kagee (2009)	PDS	15	68%	35%
Owoso et al. (2014)	mPRIME	4/5	40%	65%
Pence et al. (2012)	PHQ-9	10	27%	94%
Saal, Kagee, & Bantjes (2018)	BDI-I	20	67%	67%
Seth et al. (2015)	AUDIT-C; AUDIT-3	3; 1	95.2% (AUDIT-C); 85.3% (AUDIT-3)	77.5% (AUDIT-C); 34.1% (AUDIT-3)
Sharp et al. (2014)	SDQ-Caregiver; SDQ-Youth; SDQ-Teacher	13.5 (Caregiver); 11.5 (Youth); 11.5 (Teacher)	70% (Caregiver); 58% (Youth); 58% (Teacher)	73% (Caregiver); 51% (Youth); 56% (Teacher)
Stewart et al. (2009)	SRQ	8 (Major+Minor Depression); 8 (Major Depression)	56.4% (Major+Minor Depression); 66.2% (Major Depression)	87.6% (Major+Minor Depression); 80.8% (Major Depression)
Stewart, Umar, Tomenson, & Creed (2013)	SRQ; EPDS	6 (SRQ, Major Depression); 7 (EPDS, Major Depression); 5 (SRQ, Minor Depression);	61.8% (SRQ, Major Depression); 76.3% (EPDS, Major Depression); 83.5% (SRQ, Minor Depression); 62.5% (EPDS, Minor	76.4% (SRQ, Major Depression); 81.3% (EPDS, Major Depression); 78.7% (SRQ, Minor Depression); 85.8% (EPDS, Minor

Study	Screening Tool	Area Under Curve Cutoff	Sensitivity	Specificity
		7 (EPDS, Minor Depression)	Depression)	Depression)
Udedi, Muula, Stewart, & Pence (2019)	PHQ-9	9	64%	94%
Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016)	TSI; SSI	22 (TSI, Use/Abuse); 42 (TSI, Abuse/Dependence); 6.5 (SSI, Alcohol Use/Abuse); 14.5 (SSI, Alcohol Abuse/Dependence); 1 (SSI, Illicit Drugs Use/Abuse); 18 (SSI, Illicit Drugs Abuse/Dependence)	79% (TSI, Use/Abuse); 64% (TSI, Abuse/Dependence); 85% (SSI, Alcohol Use/Abuse); 60% (SSI, Alcohol Abuse/Dependence); 93% (SSI, Illicit Drugs Use/Abuse); 90% (SSI, Illicit Drugs Abuse/Dependence)	72% (TSI, Use/Abuse); 61% (TSI, Abuse/Dependence); 89% (SSI, Alcohol Use/Abuse); 60% (SSI, Alcohol Abuse/Dependence); 93% (SSI, Illicit Drugs Specificity); 87% (SSI, Illicit Drugs Abuse/Dependence)

V. Discussion

This in-depth review highlights three main procedures that are predominantly in use to culturally validate mental health screening tools in sub-Saharan Africa:

- Language Translation
- Terms Validation
- Method of Administration

The results illustrate that even though most studies employ these techniques, there remains great variability in the way the processes are undertaken from study to study.

Firstly, with regard to language translation, the evidence shows that a majority of studies do attempt to incorporate the ideal standard of translation that was discussed above, back translation, for a non-English speaking target population, but not all studies are using perfect back translation techniques. This review shows that back translations were conducted using many different approaches in different studies. This lack of standardization across studies diminishes the strength of some of the translations as back translation is heavily dependent on cultural or conceptual equivalence as opposed to simply linguistic equivalence. The former is achieved when those conducting the translations are not only sufficiently bilingual but also have a good knowledge of mental health terminology in both languages and a familiarity with the culture of the target population. Several studies brought in only bilingual professional translators which failed to account for the probability that working with these translators -who are language experts- produced tools with language validity but not any or minimal subject validity (Heyningen, Honikman, Tomlinson, Field, & Myer, 2018; Odenwald et al., 2007; Marquer et al., 2012; Chishinga et al., 2011; Chibanda et al., 2016) Some other studies engaged the researchers

and health workers to fulfill the translation of tools but as with professional linguists, researcher might not possess the full skillset adequate for translation and likely only possess an understanding of the content matter and not linguistic motifs (Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt, 2015; Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Blair et al., 2017; Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017; Denckla et al., 2017; Geibel et al., 2016; Gelaye et al., 2013; Hanlon et al., 2015; Shanahan, Anderson, & Mkhize, 2001; Natamba et al., 2014). This can result in tools with terms that are unknown to the culture or terms that are perceived differently from the original intent of the instruments. Despite these shortcomings in some studies, the author notes there were a few studies who utilized multiple translators, from multiple fields, who undertook the translation process independently (Udedi, Muula, Stewart, & Pence, 2019; Stewart et al., 2009; Stewart, Umar, Tomenson, & Creed, 2013). This is likely to produce tools that incorporate a broader range of cultural and linguistic content and thus are more in-depth. A good number of studies that did utilize multiple translators did so by seeking persons from different areas of expertise such as when the translations were done by a researcher and local social worker (Chorwe-Sungani & Chipps, 2018) or when it was done by local psychiatrists, researchers and professional translators. There is a proportion of studies that do not report on who conducted translations and it is possible that this could be a result of study researchers not knowing the importance of recruiting a translator who is both linguistically as well as culturally adept (Cholera et al., 2014; Flisher, Sorsdahl, & Lund, 2012; Gelaye et al., 2014; Kane, Murray, Bass, Johnson, & Bolton, 2016; Mellins et al., 2018; Muris et al., 2006; Nolan et al., 2018; Seth et al., 2015). This review emphasizes that such studies weaken the strength of their validity claims compared to studies that employ more robust methods. This review has gathered information from an adequate number of studies to infer that a majority of

studies that are conducting back-translation of mental disorder screening tools for a non-English sub-Saharan African population are doing so well and thus future studies should look to them to understand how to adequately conduct translation. Particularly those where multiple and different subject experts are utilized, because, as discussed, the broader the depth of the translators involved in translation, the more likely a tool is to be valid and researchers should always take note of this. This ensures the effectiveness of a screening tool for non-English speaking populations.

Secondly, this review has found that a substantial percentage of studies are incorporating methods to validate individual terms or items in the screening tools. This is very important as not all terms directly translate across cultures and expert consensus approaches are needed to ensure a tool is valid for its target population. Additionally, this step is important for tools that do not undergo a translation, as they are intended to be used in an English-speaking population because although people may speak the same language, their lived experiences of their culture and environment significantly affects their understanding of certain concepts. As with the language translation exercises, validation of terms was done very differently by different studies. A majority of studies employed consensus building discussions. However even this was undertaken differently by different studies. Some studies such as Chishinga et al. (2011) only sought to gain the consensus of the research team or the translators when discrepancies arose whilst others employed more thorough methods such as engaging with a wider range of experts like psychiatrists, psychologists, social workers and mental health researchers who met until each problematic item was resolved (Odenwald et al., 2007). Although the former technique is a good first step towards ensuring valid terms are incorporated, engaging a wider range of experts has its obvious strengths and should be the avenue pursued by most studies.

An alternate way of ensuring terms in screening instruments are valid that was utilized is conducting focus groups with a sample of the target population. Focus groups allow researchers to gauge the possible attitudes and behaviors towards the tools prior to implementation in a large population. Thus, ensuring the tool fully encompasses these attitudes in any revision. There was good number of studies identified in this review that conducted some form of a focus group. Although it should be noted that like the other processes discussed above, the focus groups were also quite variable in their setup from study to study. Some studies only sought focus group participants from the target population (Stewart et al., 2009; Pence et al., 2012) whereas others expanded it community key informants as well as the target population (Marquer et al., 2012; Stewart, Umar, Tomenson, & Creed, 2013). Even though the focus groups were conducted with rigor, the latter seems the better methodology to employ in most instances, in order to gain as wide a range of views as possible and thus strengthen the depth of the tool. Additionally, focus groups allow researchers to understand certain dynamics in play in a society that could impact the strength of a tool. For example, one study conducted focus groups segregated by gender and religion and it was found that certain expression of symptomatology varied between genders (Pence et al. (2012). Studies that are able to gain such information are certainly more robust as they ensure these dynamics are accounted for in the final tool bolsters true positives and reduces the likelihood of false positives.

A minority of studies incorporated pilot testing of tools prior to conducting a full study. It was found that all tools were pilot tested in a sample of the target population which can be deemed as ideal as developing internal validity is imperative to the strength of a tool when cultural validity is trying to be showcased. That is, the tool should work well within a particular culture prior to

any attempts to deem it fit to be extrapolated to other cultures. Additionally, due the great difference in cultures across sub-Saharan Africa, external validity of tools might never be fully achieved for individual tools and thus every study that uses that tool should seek to amend it to be culturally fit for that population. The reason for the limited number of pilot studies being conducted could be the cost associated with them but if and when possible, they allow researchers to gain a solid foundation prior to conducting a full study. This foundation is very important as it not only provides feedback on the items and language in the tools but very much hints at the overall feasibility of using that method of screening for mental disorders in that population.

Lastly, the method of administration (self-administered vs. interviewer administered) was a prominent feature of cultural validation of tools in sub-Saharan Africa. The chosen method of administration in this context to ensure cultural validity as many of populations in the region face high rates of illiteracy. Thus, utilizing self-administration could bias the results and either create a high proportion of false positives or false negatives. Interviewer administered studies were predominant in this region, but the choice of interviewer was variable among studies. A majority of interviews were done by the researcher or members of the research team who underwent training (Berard, Boermeester, & Viljoen, 1998; Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015) and although this is adequate in the short term, it brings up the problem of sustainability of the tool in the population following this initial validation study. Some other studies had interviews conducted by local health workers (Geibel et al., 2016; Sharp et al., 2014) which seems to account for the sustainability issues that the researcher interview studies do not. However, neither of these types of interviewers account for the possible power dynamics in play

between professionals and the study participants. This is particularly a significant factor in some places due to sub-Saharan Africa's colonial history and subsequent lingering societal disparities between those who go through western education and those who do not. Another group of studies thus opted to engage trained lay workers as interviewers (Odenwald et al., 2007; Mellins et al., 2018). Using this group as interviewers could solve the problem of sustainability of the screening tool beyond the initial validation study and the political contextual factors in sub-Saharan Africa. Despite which interviewer was enlisted, a majority of studies trained interviewers on how to properly administer the tools (Akena, Joska, Obuku, & Stein, 2013; Bertschy, Viel, & Ahyi, 1992). The strength and benefits of this is evident as it provides a standard for all interviewers and allows for the different results to be comparable. Finally, there were some studies that did not mention who conducted the interviews or the method in which the screening tool was administered (artin, Fincham, & Kagee, 2009; Seth et al., 2015; Suliman, Kaminer, Seedat, & Stein, 2005). This gap in data weakens a study's conclusion of strength of its validity exercise and more in-depth reporting supports not just individual studies but the overall body of literature on screening tools.

The studies conducted these validation techniques before conducting statistical tests on the results. Statistical tests and their results ensure there is adequate construct validity in the tools and therefore are just as important as the cultural validity of the tools, in determining their utility and reliability. However, statistical analysis is dependent on how culturally valid the tools are. If a tool produces statistically accurate results but is inherently biased, then overall caution should be observed when judging its validity. The studies in this review were robust in their statistical analysis however only about half conducted sensitivity and specificity tests. With a tool such as

the PHQ-9, most studies decided on similar cutoff values, for calculating the sensitivity and specificity (Bhana, Rathod, Selohilwe, Kathree, & Petersen, 2015; Adewuya, Ola, & Afolabi, 2006; Udedi, Muula, Stewart, & Pence, 2019).

The findings of this systematic review are important for mental health professionals and researchers alike. The use of tools that are not sufficiently validated is still widespread and this only worsens the outcomes for the sub-Saharan population. The field of mental health is still developing in sub-Saharan Africa and the future of mental health in the region needs for studies to sustain the back-translation method of translations that is already highly prevalent, better validation of terms and the involvement of multiple experts and to be cognizant of the socioeconomic status and educational characteristics of the population. Additionally, studies must ensure they conduct statistical analysis of various kinds but primarily sensitivity and specificity calculations to create and utilize unbiased and reliable tools. This is absolutely essential if we are to address the challenges of global public health and reduce the rising burden of mental disorders. This review recognizes studies such as those conducted by Pence et al. (2012) and Chibanda et al. (2016) as being the most robust in their validation and the type of studies to look to for future research.

VI. Limitations

The first limitation of this review is that it only included studies published in English. As sub-Saharan Africa has many different official and local languages, it would be important to include these articles in future reviews to get a more thorough understanding of validation techniques on the continent. Secondly, the inclusion criteria of this study only incorporated official ICD-10 disorders so as to hone the scope of this review, although the literature search highlights that broadly defined mental disorders are being studied in this population and should be incorporated in future reviews. Thirdly, this review was undertaken by a single author as part of the requirement for a master's in public health degree and is thus subject to research bias in thoroughness due to limited resources. However, searches were conducted multiple times to reduce this and the author is confident in the results obtained from this review. Fourth, studies included in this review are subject to secondary data bias and therefore I cannot validate the findings of the study. Lastly, this review excluded screening tools that were developed specifically for the sub-Saharan African context under the assumption that tools developed for populations would be valid for that population. However, a review of this could indicate some falsity in this assumption and add to the body of evidence on ways to validate screening instruments.

VII. Conclusion

As the burden of mental disorders continues to rise in sub-Saharan Africa, it is imperative for mental health care to be more diligently focused on and this starts with screening and diagnosing individuals. However, for the screening tools to capture true disorders they must be sound for the population and incorporate their cultural understanding of mental health.

This review sought to bridge a current gap in literature and to discern what is currently being done to ensure this cultural and construct validity in sub-Saharan Africa. This review found that many studies attempted to ensure cultural validity in their tools using three main procedures. First, translation of tools for a non-English speaking population was used in a high number of studies. Although the methods of translation varied, it is evident that a more robust undertaking is more favored such as that undertaken by studies that conducted back translation with both bilingual translators as well as conceptual experts. Secondly, the review found validation of items in the instrument to be a widely conducted exercise in cultural validation for these studies. This is imperative for dispelling any disparities between the population's understanding of mental health and the original instrument's intent. As with translation various techniques were chosen, namely: conducting consensus building exercises, conducting focus groups and/or conducting pilot tests. The use of all three techniques, if possible, certainly bolsters the validity of the instrument in the population especially if these are conducted using a wide range of experts and they sample in an in-depth population pool. Third, the review identified two different methods of administering the tools: self-administered vs. interviewer administered. The chosen method can impact the ability of study participants to fully understand the concepts being conveyed by the study and with high rates of illiteracy amongst many sub-Saharan African populations, the method chosen could have a great impact on the true and false

positive results attained by screening tools. Last but not least, many reviews conducted analytical tests to ensure the external and internal reliability of tools was evident in the population and this should be done for all studies.

Translating tools, ensuring conceptual equivalence of their individual items, administering the tools and then analyzing the results in a manner adequate for the population are good first steps in ensuring validity. Nevertheless, more techniques can be employed, and current ones enhanced to fully ensure the attitudes and behaviors, with regards to mental health, of a particular population are fully incorporated in screening tools. This is imperative to curb the rising burden of mental disorders in sub-Saharan Africa.

Appendixes

Appendix A

Study	Disorder(s)	Country	Population	Tool	Criterion	Language Translation	Terms Validation Methods	Terms Validation Results	Other
Adewuya, Ola, & Afolabi (2006)	Depression	Nigeria	University Students	Patient health questionnaire (PHQ-9) and Beck Depression Inventory (BDI)	Mini International Neuropsychiatric Interview (MINI)	The interviews were conducted in English, which is the official and the most common form of communication in Nigeria	-		-
Akena, Joska, Obuku, & Stein (2013)	Depression	Uganda	HIV+ Adults	Patient Health Questionnaire-9 (PHQ-9), Centre for Epidemiological Surveys for Depression (CES-D), and the Kessler-10 (K-10)	Mini International Neuropsychiatric Interview (MINI)	Not mentioned	-	-	Two triage nurses were trained by a psychiatrist administered the screening instruments
Andersen et al. (2011)	Mood and Anxiety Disorders	South Africa	Adults	K6 and K10 scales	Composite International Diagnostic Interview (CIDI)	Yes, to Afrikaans, Zulu,Xhosa, Northern Sotho, and Tswana	-	Discrepancies that emerged from the back-translation were decided on	-

								by consensus of a team of experts	
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	Antenatal depression	Cote d'Ivoire and Ghana	3rd Trimester Woman	Patient health questionnaire (PHQ-9)	-	Yes, to French (CDI) and Twi (Ghana)	Both language versions were reviewed by a committee to ensure their equivalence; Questionnaires were pilot-tested to ensure the comprehensibility of the items.	N/A	All questionnaires were administered as interviews by three trained psychologists in each hospital; All interviewers were trained by psychiatrists/psychologists both from Germany and the respective country; DIF analyses within the IRT framework (bias amongst understanding between patients); Questionnaires were pilot-tested to ensure their suitability in the setting
Berard, Boermeester, & Viljoen (1998)	Depressive disorders	South Africa	Out-patient cancer patients (new and follow up)	Hospital Anxiety and Depression Scale (HADS) and Beck Depression Inventory	Clinical Interview	Not mentioned	-	-	A trained research assistant attended weekly breastcancer, head and neck cancer, and lymphoma clinics over a period of

	(BDI)	approximately 2 years for patient selection; Illiterate patients were excluded; Patients completed the scales in the waiting room; Researchers assisted by reading out certain questions and asking for clarification if needed; A psychiatrist, clinical psychologist, and health psychology intern (who is also a qualified psychiatric nurse) conducted the psychiatric interviews; The interviewing clinicians met for several training sessions wherein they agreed to follow a structured interviewing format; Interviews took place in the nurses' station or temporarily

									unused consulting rooms
Bertschy, Viel, & Ahyi (1992)	Depression	Benin	Patients at the neuropsychiatric out-patient clinic of the National University Center of Cotonou	Comprehensive Psychopathological Rating Scale (CPRS)	-	French version was used; previously translated	Proposed a modified version of the DSM III criteria for major depression; During the four first months the author explored the feasibility of the clinical rating with the CPRS to become familiar with the way in which Beninese people express what they feel, to evaluate the local meaning of certain words or the understanding of certain concepts	Preserved the global structure of the DSM III criteria; Taken into account specificity of depression in Black Africa: depression without dysphoric mood, ideas of persecution and somatic complaints; Authors went progressively deeper into interviews with a kind of cross-cultural back translation: every time a patient seemed to report a symptom, spontaneou	French psychiatrist of senior registrar level spent a period of 14 months in Benin and trained in the use of clinical rating instruments

								sly or after having been asked, we asked him to give us a concrete description using several examples; At the end of this prestudy stage, authors started to make complete ratings with the CPRS	
Betancourt et al. (2009)	Mood disorders (two tam, par and kumu), General anxiety disorder (ma lwor) and Conduct disorder (kwo maraco)	Uganda	Acholi youth aged 14-17 years and their caregivers in camps	Acholi Psychosocial Assessment Instrument (APAI)	-	Originally in Acholi Luo	45 free listing interviews with local youth and adults and 57 in-depth interviews with local key informants to identify and describe the important problems of youth in these camps	Qualitative study identified eight items indicative of prosocial behavior among youth in this setting	All interviewers and supervisors received 3 days of training in research ethics, interviewing techniques and questionnaire administration

							for internally displaced people (IDPs)		
Betancourt, Yang, Bolton, & Normand (2014)	Psychosocial adjustment disorder	Uganda	Youth aged 14-17 living in Unyama and Awer IDP camps	African Youth Psychosocial Assessment (AYPA)	-	Originally in Acholi Luo	-	-	Trained local interviewers administered the full APAI to all youth referred for screening; All assessments were conducted in private following informed consent and youth assent
Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015)	Depression	South Africa	Chronic care patients	Patient Health Questionnaire (PHQ-9)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Yes, to seTswana	-	-	Bilingual, trained research assistants orally administered the PHQ-9 screening tool, and entered the participant's responses in a questionnaire application programmed on a mobile handheld device and were supervised by mid-level psychological counsellors with 4-year Bachelor's degrees in psychology who were fluent in seTswana; 1 SCID psychologists

									was fluent in seTswana and english and the other was assisted by a seTswana speaking mid level trained psychological counselor; Piloting occured and responses were adapted to improve respondent understanding
Blair et al. (2017)	Alcohol Misuse, PTSD and Depression	Uganda	13-49-year olds	AUDIT, Hopkins Symptom Checklist-25 (HSCL-25) and Harvard Trauma Questionnaire (HTQ) Parts I and IV	-	Yes, to Acholi Luo	Translated by experienced Acholi researchers and tested in a pilot community	-	-
Carney, Myers, & Louw (2016)	Substance mis-use	South Africa	Adolescents	CRAFFT, Global Appraisal of Individual Needs-Short Screener (GAIN-SS) substance use subscale and Personal Experience Screening Questionnair	-	No, but an Afrikaans and isiXhosa project staff member was present	-	-	Self completion but assistant was available if needed

				e (PESQ)					
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Maltreatment, PTSD and Depression	Burundi	Adolescents aged 8-21 years living with an adult parent	Childhood Trauma Questionnaire (CTQ), Depression Self-Rating Scale (DSRS) and Child PTSD Symptom Scale (CPSS)	-	Yes, from French to Kirundi	Each item in Kirundi was followed by extensive discussions in the group until all members reached an agreement that the items would have semantic, conceptual and content equivalence across different socio-economic strata of the society; Like the CTQ, all the depression and PTSD items were discussed by the group of bilingual professionals, and slightly adapted when the	The team, after extensive deliberation, added the word 'kenshi' (referring to 'often' in English) to item 24 comprising the subscale of sexual abuse	Data collection was carried out by 8 male and 8 female interviewers bilingual in French and Kirundi with a University degree and prior experience with survey methods in Burundi, who received a half-day of additional training on the methodology and measures used in the present study; Care was taken to have equal number of interviewers whose ethnic background was Tutsi or Hutu (the two major ethnic groups in Burundi); No information of the ethnic background of the participants was collected and no attempt was made to match the ethnic background of interviewers and

							group thought it to be necessary.		interviewees; Verbal informed consent from all adult participants and verbal assent from adolescents was collected in place of written consent
Chibanda et al. (2016)	Depression and Anxiety disorders	Zimba bwe	HIV Patients	The Shona Symptom Questionnaire for common mental disorders (SSQ-14), PHQ-2, the Patient Health Questionnaire for depression (PHQ-9), and the Generalized Anxiety Disorder questionnaire (GAD-7)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Yes, the PHQ-9 and GAD, to Shona	Draft Shona version was reviewed by a team focused on ensuring contextual equivalence to the original versions based on their understanding and use of local terms for mental distress including kufungisisa (thinking too much), kusuwa (to feel sad), kusuwisisa (to feel very sad) and mwoyo unorwadza (painful	A committee of the first author with the clinical psychologist and counselor examined both the original and back translated versions and resolved any discrepancies by consensus	SSQ-14: The Shona Symptom Questionnaire was developed and validated in Zimbabwe and includes local idioms of emotional distress including 'thinking too much'; The research assistants were trained in data collection methods using the socio-demographic forms and the screening tools; The psychiatrists were trained in the use of the SCID through a discussion forum led by DC which involved going through the diagnostic criteria, building

							heart)		consensus on how to manage clinically severe cases during the validation, and procedures for ensuring fidelity; DC observed the psychiatrists during role-play and the pilot validation where each psychiatrist used the SCID to interview 4 patients
Chishinga et al. (2011)	Depressive and alcohol use disorders	Zambia	TB and HIV adults	Center for Epidemiological Studies Depression scale (CES-D) and Alcohol Use Disorders Identification Test (AUDIT)	Mini International Neuropsychiatric Interview (MINI)	Yes, to Bemba, Nyanja, Tonga and Lozi	All translated versions of the questionnaires were discussed by the research team comprising of members who were fluent in the native languages; Instrument was tested before use in the field	Discrepancies in conceptual and semantic equivalence was resolved through informal committee consensus	Screening interview was conducted by the trained lay research assistants who had been trained for 2 days
Cholera et al. (2014)	Depression	South Africa	High HIV burden, low literacy	Patient Health Questionnaire-9 (PHQ-9)	Mini International Neuropsyc	Translation method not mentioned but isiZulu, isiXhosa, seSotho, seTswana versions used	-	Questions about substance abuse and	Administered by trained interviewers; The MINI interview

			PHC population ; PHC Patients		hiatric Interview (MINI)			knowledge of HIV status (from prior testing experiences) were also included in the questionnaire.	was conducted by health care professionals trained in use of the instrument; The MINI interview used an interpreter when needed.
Chorwe-Sungani & Chipps (2018)	Antenatal Depression	Malawi	Pregnant women	Edinburgh Postnatal Depression Scale (EPDS), Hopkins Symptoms Checklist-15 (HSCL-15), Self-Reporting Questionnaire (SRQ) and Pregnancy Risk Questionnaire (PRQ), 3-item screener for depression.	-	Yes, only the HSCL-15, PRQ and 3-item screener for depression; EPDS and SRQ were previously translated	-	-	The first author and a second research assistant read the questions and recorded the answers on behalf of respondents.
Denckla et al. (2017)	Anxiety Problems, Affective Problems, Somatic Problems, Attention Deficit/Hyper	Kenya	School aged children 10-18	Ndetei-Othieno-Kathuku Scale	Youth Self Report DSM-IV oriented scales	Yes, to Kiswahili and Kikamba	Discrepancies in translation were reconciled by an expert committee of at least	Panel members suggested alternatives and agreed on acceptable solutions	-

	activity Problems, Oppositional Defiant Problems, Conduct Problems, Obsessive-Compulsive Problems, and Post-Traumatic Stress Problems						six multilingual experts in English, Kiswahili, and Kikamba; Translated scale was administered to a pilot sample for accuracy		
Flisher, Sorsdahl, & Lund (2012)	Any anxiety disorder (including agoraphobia), agoraphobia, major depressive disorder (MDD), attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD).	South Africa	Youth	Diagnostic Interview Schedule for Children (DISC-IV)	-	Translation method not mentioned but Xhosa version was used	Additional changes were made in the process of training the fieldworkers	-	Research assistants without previous clinical training were responsible for obtaining data from interviews with parents/caregivers and youth and worked in pairs, one interviewed the parent and the other interviewed the youth.
Francis, Helander, Kapiga, Weiss, & Grosskurth (2015)	Alcohol misuse	Tanzania	College youths and casual laborers, aged 18-24	Alcohol Use Disorders Identification Test (AUDIT)	Mini-International Neuropsychiatric Interview (MINI)	Not mentioned	-	-	-
Geibel et al.,	8 syndromes	Ethiopi	Youth	Youth Self-	-	Yes, to Amharic	-	-	Modified to be

(2016)	measuring affective disorders, anxiety problems, somatic problems, ADHD, oppositional defiant disorder and conduct disorders	a	aged 15-18	Report (YSR)					administered by trained nurses; A committee of psychiatrists, psychologists, public health professionals, service providers, and social science researchers was convened to review and discuss screening tools previously validated outside of Ethiopia
Gelaye et al. (2013)	Major depressive disorder	Ethiopia	Outpatient adults	Patient Health Questionnaire-9 (PHQ-9) questionnaire	Clinical Assessment in Neuropsychiatry (SCAN)	Yes, to Amharic	-	-	-
Gelaye et al. (2014)	Major depressive disorder	Ethiopia	Outpatient adults	Patient Health Questionnaire-9 (PHQ-9)	Composite International Diagnostic Interview (CIDI); Clinical Assessment in Neuropsychiatry (SCAN)	Translation method not mentioned but Amharic version was used	-	-	-
Habtamu et al. (2017)	Schizophrenia, bipolar disorder or major depressive	Ethiopia	People with SMD presenting with new onset of	36 item patient version of the WHODAS	Brief Psychiatric Rating Scale (BPRS-E)	Amharic version was used; previously translated	Administered the scale to a sample of people with SMD	Iterative adjustments were made to the	Data were collected by lay interviewers; BPRS-E administration by

	disorder with psychotic features	illness or relapse of existing illness	2.0			and their caregivers (participants were from rural areas and had primarily level educational) for cognitive interviewing to identify any difficulty with understanding of individual items and response categories, as well as acceptability and burden of the whole scale; Expert panel composed of psychiatrists, psychologists, social workers and mental health researchers	Amharic translations, including the addition of examples to items asking about broad and abstract experiences, and replacing less relevant and uncommon concepts to the setting with equivalent but local experiences; For the item "sexual activities" the Amharic translation was changed to ask about romantic relationships.	a psychiatrist

							suggested how each problematic item should be rephrased to be more easily understood by the respondents, while retaining semantic equivalence with the original scale.		
Hanlon et al. (2015)	Major Depressive Disorders	Ethiopia	PHC Patients	Patient Health Questionnaire (PHQ-9/PHQ-2), the 20-item Self-Reporting Questionnaire (SRQ-20) and the Kessler scales (K6/K10)	Mini International Neuropsychiatric Interview (MINI)	Yes, PHQ-9, to Amharic; Semantic, technical and content validity already established for the Amharic versions of the SRQ-20 and K10	The final version was obtained by expert consensus	-	Modifications from a self-completed to an interviewer-administered questionnaire; The lay data collectors were trained for four days by senior mental health researchers, followed by one day of observed pilot interviews; Three psychiatric nurses were trained by an Ethiopian psychiatrist for two days in administration of

									the criterion measure (MINI); SRQ items are single-clause questions which require a yes/no response and are easily administered in an interview format; A culturally adapted version of SRQ ('SRQ-F') lengthened the scale with little improvement in psychometric properties; A question on irritability was added to the MINI, based on previous qualitative work and clinical experience of depression in the setting
January & Chimbari (2018)	Postnatal depression	Zimbabwe	Postnatal women	Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire (PHQ-9) and Center for	Diagnostic and Statistical Manual for Mental Disorders (DSM-5)	Shona version was used; previously translated	-	-	-

				Epidemiological Studies-Depression (CES-D)					
Kagee, Bantjes, Saal, & Sefatsa (2019)	Post-Traumatic Stress Disorder	South Africa	People seeking HIV tests	Posttraumatic Symptom Scale, Self-report version (PSS-SR)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	No, most participants were able to converse in English	Consulted with one of the SCID developers, Dr Michael First; Conducted a pilot study to ensure that participants found the SCID items comprehensible and feasible	SCID questions were adjusted so that they corresponded to the DSM5 diagnostic criteria; Adapted the language of the instrument accordingly so that it conformed to the criteria for PTSD as delineated in the latest version of the DSM	Data collectors received training in how to administer the interviews
Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride (2018)	Depression	Kenya	Adolescents aged 13-18	Child Depression Inventory (CDI)	-	Yes, to Kiswahili	-	-	Instruments were administered by bilingual researchers in English but students had the option to respond in either language
Kane, Murray,	Substance misuse	Zambia	Orphans and	Audio Computer	Youth Self Report	Translation method not mentioned but Nyanja	All questions	Corrections were	ACASI allows respondents to

Bass, Johnson, & Bolton (2016)			vulnerable children (OVC) ages 13-17	Assisted Self-Interviewing (ACASI) system for the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST), version 3.1	DSM-IV oriented scales	and Bemba versions were used; ACASI was recorded in each of the three study languages, English, Bemba, and Nyanja and participants were able to complete the interview in the language of their choice	were then reviewed by community groups of adolescents from our target population for conceptual understanding and translation accuracy	made as needed	complete the questionnaire using a laptop computer and headphones, reading on their own and hearing a recorded voice speak the questions and response options; Participants were provided a brief training on how to use the laptops; Research staff were stationed nearby to answer questions and provide assistance, if necessary.
Marquer et al. (2012)	Depression, phobia, anxiety, regression, psychosomatic complaints, and post-traumatic disorder.	Niger	Children 3-6 years without a mental disorder, mental retardation, development disorder, and/or psychosis	Psychological Screening for Young Children aged 3 to 6 (PSYCAa3-6)	Clinical Interview	Yes, to Hausa	The translated tool was administered to a small group of caregivers in a pilot phase for acceptance, adequacy and applicability; Focus group discussions with	A list containing misunderstood items developed and an alternate proposed translation was made, the final version was fixed by the two translator's consensus taking into	Interviewers were fluent in Hausa and French and were trained over a three-day period on how to administer the questionnaire

							community key informants, individual interviews and expert meetings were conducted to elucidate understanding of the beliefs and perceptions about child development and expectations of the outcome of treatment; Consultations with psychiatrists and anthropologists and the health care staff of hospitals to define developmental milestones and child psychopathology were also organized.	account differences between spoken and written Hausa;	
Martin,	PTSD	South	HIV+	17-item PDS	Composite	Yes, clinical interview	-	-	CIDI module is

Fincham, & Kagee (2009)		Africa	Adults diagnosed within the year		International Diagnostic Interview (CIDI) Diagnostic Interview (CIDI)	was conducted in English or Afrikaans; Self-report measures were in English, Afrikaans, and Xhosa			introduced to the participant by the interviewer with the following statement: "I am now going to ask you some questions around your reactions to hearing that you were HIV-positive and to being HIV-positive", followed by a statement that reflects the DSM-IV criteria for PTSD
Mellins et al. (2017)	Major depression syndrome (MDS), other depression (ODS), panic, generalized anxiety, post-traumatic-stress-disorder (PTSD), and SUD	South Africa	Female caregivers	Client Diagnostic Questionnaire (CDQ) lay version	Client Diagnostic Questionnaire (CDQ) clinical version	Yes, to isiZulu	US and SA Principal Investigators (PI) reviewed discrepancies for meaning and correct diagnostic translation	Corrections were made as needed	Lay counselors were trained and supervised by experienced psychologists and physicians and there were ongoing weekly supervision included training reinforcement, fidelity checking, and support for handling difficult interviews and distressed caregivers
Mellins et al. (2018)	Behavioral difficulties and pro-social strengths	South Africa	Children 4-8 years old	Strengths and Difficulties Questionnair	-	Translation method not mentioned but isiZulu version was used; English also available	Discussions to resolve discrepancies and	Corrections were made as needed	Administered by lay fieldworkers who were native speakers of

				e (SDQ)			reviewed with the SDQ author; Focus groups were conducted with SA native isiZulu speakers;		isiZulu and bilingual in English and were trained and supervised by a SA child psychologist and a medical doctor.
Morojele et al. (2016)	Alcohol Misuse	South Africa	HIV+ patient on ART for atleats 3 months	AUDIT-C, AUDIT-3, AUDIT-QF, AUDIT-PF, AUDIT-4, M-FAS	AUDIT-10	Not mentioned	-	-	Screening was conducted by trained interviewers in a private room at the clinic in two steps; 10 AUDIT questions were asked verbatim by the trained interviewers using a pen-and-paper approach
Muris et al. (2006)	Anxiety disorders	South Africa	School aged youth	SCARED and EMBU-C	-	Translation method not mentioned but Afrikaans version was used; English also available	-	-	Participants were tested in their classrooms at school.
Natamba et al. (2014)	Major depressive disorder	Ugand a	Pregnant women with HIV	Center for Epidemiologic Studies Depression (CES-D) scale	Mini International Neuropsychiatric Interview (MINI)	Yes, to Acholi and Langi	Conceptual and semantic equivalence were resolved through discussion involving all the translators,	Corrections were made as needed	CES-D was administered by members of the PreNAPs study team (either a nutritionist, midwife, or general nurse)

							the research assistants, and the GRRH psychiatrist		
Nolan et al. (2018)	Depression	Tanzania	HIV+ adults	PHQ-2 and WHO-5	Patient Health Questionnaire-9	Translation method not mentioned but Kiswahili version was used	-	-	Questionnaires were administered by Tanzanian research assistants fluent in Kiswahili.
Odenwald et al. (2007)	PTSD	Somaliland	Former combatants	Posttraumatic Diagnostic Scale (PDS)	Composite International Diagnostic Interview (CIDI); Self-Report Questionnaire (SRQ-20)	Yes, to Somali	The translation and cultural adaptation were carried out by groups of local bilingual and international experts, who had received education in trauma-related concepts; Group discussions and consultations with external specialists were dedicated to culturally	Adapted to the Somali language, culture, and Islamic religion (Somali-PDS); Repeated back translations as needed	Six local interviewers underwent a 10-day training; Self-report scoring of the PDS was adapted to an 'assisted self-report'.

							specific meanings of items and typical experiences in Somalia		
Ogle, Koen, & Niehaus (2018)	Depression and anxiety disorders (PTSD, Panic disorder, Agoraphobia, Phobia and Generalized Anxiety)	South Africa	Adults with hypertension and diabetes	visual screening tool for anxiety disorders and depression (VISTAD)	Mini International Neuropsychiatric Interview (MINI)	Not mentioned	-	-	-
Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	Post stroke depression	Nigeria	Stroke survivors	26-items Health Related Quality of Life in Stroke Patients (HRQOLISP-26)	Hospital Anxiety and Depression Scale (HADS-D)	Not mentioned	Consultant psychiatrist independently examined each of the Psycho-emotional items and determined whether the scale contained the minimum number of items required for the diagnostic assessment of depression specified in the ICD 10.	-	-
Opakunle,	Suicide risk	Nigeria	Outpatient	Obsessive-	Mini-	Not mentioned	-	-	-

Aloba, Akinsulore, Opakunle, & Fatoye (2018)			s with Schizophrenia	Compulsive Inventory-Revised (OCI-R) and Positive and Negative Syndrome Scale (PANSS)	International Neuropsychiatric Interview (MINI)				
Owoso et al. (2014)	Psychotic Disorder	Kenya	Youth 14-19	PRIME screen (mPRIME)	Psychosis-Risk Syndromes (SIPS)	No, conducted in English	-	Ten of the twelve mPRIME questions were identical to the original PRIME screen questionnaire with the only differences being replacement of items 9 and 12	Research assistants underwent an extensive 2-day training in the SIPS background and administration by a psychiatrist (D.M.) with significant prior experience in conducting the SIPS interview, formal certified training in the SIPS and PRIME by the developers of the assessments at Yale University, and extensive clinical and research experience in psychotic disorders
Pence et al. (2012)	Depression	Cameroon	HIV Patients on ART	Patient Health Questionnaire-9 (PHQ-9)	Composite International Diagnostic	No, more than 95% of clinic attendees can communicate effectively in English	Four focus groups were conducted with	Analysis suggested minor wording	Study staff trained in qualitative methods led the

					Instrument (CIDI)		hospital patients and family members, groups were separated by gender and religion (Christian and non-Christian) and were Audio-recorded and transcribed for content analysis; Optional clarifications (that were read only if the original question was not understood)	changes to two of the questions and optional clarifications (that were read only if the original question was not understood) for six additional questions.	focus groups; PHQ-9 was read to all participants and responses were recorded on the form by the study team member; CIDI interviewer was a health care professional who received formal CIDI training from a WHO-certified trainer
Saal, Kagee, & Bantjes (2018)	Major Depressive Disorders	South Africa	HIV test seekers	Beck Depression Inventory	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	No, study eligibility was understanding English	Prior to the commencement of the study, we conducted a pilot study to determine the comprehensibility and feasibility of the SCID among the	The results of the pilot study showed that the SCID was suitable for use among individuals seeking HIV testing in South Africa	The assessment instruments were administered by five psychology postgraduate students who had received extensive training and supervision from two of the authors.

							population of interest;		Participants were free to not answer any questions that made them uncomfortable and to stop the interview at any time; Surveys were administered in a private location of participants choosing
Seth et al. (2015)	Alcohol use disorder	Namibia	Adults	AUDIT, AUDIT-C and AUDIT-3	AUDIT-10	Translation method not mentioned but English, Oshiwambo, or Afrikaans versions used	-	-	
Shanahan, Anderson, & Mkhize (2001)	Psychological distress measured in terms of nine primary symptom dimensions: Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism	South Africa	Farm workers	SCL-90-R	-	Yes, to Zulu	The translators were instructed to rate each item according to the difficulty encountered in achieving what they regarded as an equivalent translation; All 90 items, as well as the instructions, response categories, and example, were	Six items (numbers 30,47,57,64,65, and 86) were identified by the committee as requiring decentering in order to facilitate translation into Zulu; Based on committee discussions, Five SCL-90-R items (numbers 5, 9, 12,62,78) were	Pretest was conducted; The Zulu SCL-90-R was administered verbally by trained Zulu research assistants who were instructed to spend some time discussing the SCL-90-R to ensure that each participant understood what was required of him before commencing with the testing; The response categories were repeated for each item.

	reviewed; Particular attention was paid to items which had (a) been rated by the translators as presenting problems in achieving equivalent translation (ratings of 2 or 3), and/or (b) resulted in back-translations which were nonequivalent to the original English items; A committee, consisting of the three Zulu psychologists and the psychiatric nurse, met to consider two back-translated Zulu versions of the SCL-	identified as possibly requiring revision in order to make the items more comprehensible or less offensive to respondents and were revised to the cultural context; For the final tool, 48 items from the first translation were selected, 35 from the second translation and 8 items were rewritten to include a combination of elements from both translations

							90-R and each item was considered individually in the committee, the pros and cons of different Zulu versions of the items being openly debated over several meetings, totaling more than eight hours of committee time.		
Sharp et al. (2014)	Anxiety disorders, affective disorders, Oppositional-Defiant Disorder (ODD), Conduct Disorder (CD) and ADHD	South Africa	Orphaned and vulnerable children aged 7-11 years	Strengths and Difficulties Questionnaire (SDQ)	NIMH Diagnostic Interview Schedule for Children (CDISC-IV)	Yes, to Sesotho	Native Sesotho translators met following individual translations; All translators met to agree on the final version	-	During the training of community mental health workers who administered the CDISC-IV, emphasis was placed on the following cultural theme: the structure of the CDISC-IV, its computerized nature, Americanisms,

problems in interpretation due to the adversity children live under, language problems, the effect of rural settings and education level, and cultural norms regarding psychiatric symptoms, gender, the experience of time, the expression of emotion, and family structure; Training involved 4 days, including teaching on the philosophical underpinnings of the CDISC-IV, its development, its installation, and its administration; Each of the relevant modules was discussed and a demonstration of the use of the CDISC-IV was given; Each of the fieldworkers received their own

									computerized copy of the CDISC-IV as well as the CDISC-IV manual; Home-based assessments in a quiet room; Study staff assisted research participants by reading questions and response options to participants;
Stewart et al. (2009)	Maternal Depression	Malawi	Mothers	Self-Reporting Questionnaire (SRQ)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Yes, to Chichewa	The 4 translations were reviewed together so variations between the translations of each item were discussed; 3 focus groups were held, each with 4 to 6 participants who were invited to discuss and comment upon the SRQ translation; Pilot SRQ	Consensus reached based on meeting; Further modifications were made on the basis of the back-translations; After each focus group further modification was made to the translation based on the comments of the	EU and MW were trained in administration of the SCID, and conducted over 30 pilot interviews; SCID interviews took place in a private room; A set of standardized brief explanations/examples for SRQ were developed for clarification and interviewers were instructed to read the explanation, explain that the respondent should give the answer that most closely (if not

			was administered to 12 mothers/guardians of infants admitted to a nutritional rehabilitation unit and problematic items were discussed with the women in order to check how they had understood the item	participants; The interviewers noted any items about which the pilot subjects asked for clarification; Experience from the piloting was discussed by the group and final version was agreed; A number of items presented difficulties in translation, as some of the English terms did not have an exact equivalent in Chichewa and in these cases either the	exactly) describes their experience, and then read the original item once more

								most closely matching word was used, or two words that between them fully captured the English meaning were included, each item was prefaced with "In the last 4 weeks…" and in one case the order of items was changed to avoid misinterpretation	
Stewart, Umar, Tomenson, & Creed (2013)	Antenatal depression	Malawi	Pregnant Woman	Edinburgh Postnatal Depression Scale (EPDS); Self Reporting Questionnaire (SRQ)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Yes, to Chichewa; SRQ and SCID previously translated	RS advised on the concepts captured by the original English wording of each item to guide the choice of Chichewa expression;	Translation followed the method used in the development of the SRQ in Chichewa previously described; The choice of	Only those whose preferred language was Chichewa were included; SCID interviews took place in a private room near to the antenatal clinic

						Discussion was held with nurses from the antenatal clinic who were invited to comment upon the EPDS translation; Piloted to a number of women attending the antenatal clinic	Chichewa expression was based on RS advising; Modifications were made on the basis of the back-translation; Further modification was made to the translation based on nurses' comments; Problematic items were discussed with focus group women in order to check understanding and final version was agreed; The use of a visual prompt card occured to

							facilitate the responses.	
Suliman, Kaminer, Seedat, & Stein (2005)	PTSD	South Africa	Youth	Child and Adolescent Trauma Survey (CATS)	Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL)	Not mentioned	-	-	Administered without parents
Tomita, Kandolo, Susser, & Burns (2016)	Depression	South Africa	non-south African refugees, asylum seekers and migrants	16-item, four-point Likert scale self-reported South African version of the Quick Inventory of Depressive Symptomatology (QIDS)	Clinical Interview	No, participants had the ability to read/speak English	-	-	Interviews were conducted at the social service organization by a trained master's degree volunteer who had refugee status in South Africa
Udedi, Muula, Stewart, & Pence (2019)	Depression	Malawi	Type-2 diabetics	Patient Health Questionnaire-9 (PHQ-9)	Diagnostic interview (the Structured Clinical Interview for DSM-IV) (SCID)	Yes, to Chichewa	Evaluation of the translated tools by the principal investigator, two mental	Consensus translation arrived at after evaluation; Pretesting helped to identify	Research assistants with bachelor's degree had 3 days of training in administration of the research tools and spoke both

							health professionals and two health promotion officers with extensive expertise in developing health communication tools; Pretested the Chichewa translated version of the PHQ 9 on a convenience small sample of 15 participants where participants were probed about their perceived interpretation of the constructs	any challenges respondents might have with the translation and unclear Chichewa terms were modified to include terms that are more commonly used and understandable by the participants	English and Chichewa.
van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Antenatal depression or anxiety disorders	South Africa	Pregnant Women	The Edinburgh Postnatal Depression Scale (EPDS), The	Mini International Neuropsychiatric Interview	Yes, to isiXhosa and Afrikaans; Diagnostic interview conducted in Afrikaans, with isiXhosa interpretation if needed by	-	-	All five screens were verbally administered by a field worker with an honours degree in

				Patient Health Questionnaire (PHQ-9), The Kessler Psychological Distress scale (K10) and a shortened 6-item version (K6), The Whooley questions and The two-item Generalised Anxiety Disorder scale (GAD-2).	(MINI)				psychology and four years of experience in clinical settings
Westhuizen, Wyatt, Williams, Stein, & Sorsdahl (2016)	Alcohol and substance use disorders	South Africa	EC Patients	Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)	Mini International Neuropsychiatric Interview (MINI)	Not mentioned	-	-	The field workers (bachelor-level university graduates) who recruited and interviewed some patients had been trained in the use of the questionnaires, apart from the MINI but the majority of the patients were recruited and screened by Dr van der Westhuizen (medical doctor),

									who had been trained in the use of the MINI
Woldetensay et al. (2018)	Depression	Ethiopi a	Pregnant women	9-item Patient Health Questionnair e (PHQ-9)	Mini Internation al Neuropsyc hiatric Interview (MINI)	Yes, to Wellega and Tulama Afaan Oromo dialects	A reconciliati on meeting was conducted to identify any areas presenting linguistic problems and to assess the participants ' level of understandi ng	All items were revised for comprehen sion (meaning and question objectives) , informatio n retrieval (type of informatio n and recall strategy), decision process (sensitivity and social desirability), and adequacy of response options)	To simplify administration as an interview rather than a self-administered questionnaire, each statement was converted into question form

Appendix B

Study	Statistical Analysis	Statistical Results
Adewuya, Ola, & Afolabi (2006)	Sensitivity; Specificity; Predictive values; Overall Correct Classification (OCC) rates; Cronbach's alpha coefficient; Spearman correlation; ROC Curve	Spearman's correlation: PHQ-9= 0.560, BDI= 0.422; Cronbach's alpha: 0.85; PHQ-9 (Major+Minor Depression): Sensitivity= 0.897, Specificity= 0.989, PPV= 0.875, NPV= 0.981, OCC= 0.973; PHQ-9 (Major Depression): Sensitivity= 0.923, Specificity= 0.984, PPV= 0.600, NPV= 0.998, OCC= 0.983; Concurrent validity: r = 0.67
Akena, Joska, Obuku, & Stein (2013)	Receiver operating curve (ROC); Areas under the curve (AUC)	PHQ-9: AUC score= 0.964, Sensitivity= 0.91, Specificity= 0.81; The PHQ-2: AUC score= 0.82, Sensitivity= 0.83, Specificity= 0.70; CES-D: AUC score= 0.944 CI (0.89 0.99). At the optimal score of 18, the CES-D was 0.88 sensitive, with a specificity of 0.81; K-10: AUC score= 0.824, Sensitivity= 0.83, Specificity= 0.72; K-6: AUC score= 0.81, sensitivity= 0.77, specificity= 0.67
Andersen et al. (2011)	ROC analyses; Likelihood ratios (LRs); Cronbach alphas	For any 12-month depression/anxiety disorder AUC (K10)= 0.73, AUC (K6)= 0.72; Sensitivity (K10)= 70, Sensitivity (K6)= 70.2 ; Specificity (K10)= 66.7, Specificity (K6)= 67.9; PPV (K10)= 23, PPV (K6)= 20.7; LR+ (K10)= 2.1, LR+ (K6)= 1.84; LR- (K10)= 0.45, LR- (K6)= 0.48;
Barthel, Barkmann, Ehrhardt, Schoppen, & Bindt (2015)	CCT: Item intercorrelations; Cronbach's alpha; Factorial validity: CFA; IRT: Item intereliability; Response categories; Sample and Item Fit; Convergent Validity	Cronbach's α: 0.65 in CDI, 0.68 in Ghana; Item intercorrelations: CDI: r=0.00 - 0.32, Ghana: r=0.02 - 0.40; Item reliability: 0.99 for both CDI and Ghana; Person reliability: 0.53 for CDI, 0.55 for Ghana; DIF Effects (interviewer): 1, 2, 6, 8 and 9 in CDI, 2, 3, 7 and 9 in Ghana; Covergent validity: CDI: r=.42, Ghana: r=.39
Berard, Boermeester, & Viljoen (1998)	Sensitivity; Specificity; PPV; Misclassification rates	BDI: Sensitivity: 85%, Specificity: 95%, PPV: 82% and Misclassification rates: 7%; HADS: Sensitivity: 71%, Specificity: 95%, PPV: 79% and Misclassification rates: 10%; Assessment Battery: Sensitivity: 95%, Specificity: 91%, PPV: 74% and Misclassification rates: 8%

Bertschy, Viel, & Ahyi (1992)	Principal component analysis	Principal component analysis variance: Mood alteration and motor slowness= 17.4%, Psychic slowness= 10.3%, anxiety= 6.3%, physical asthenia= 5.8% and persecution= 5.1%
Betancourt et al. (2009)	Internal consistency: Spearman-Brown "split half" and Cronbach's alpha coefficients; Test-retest and inter-rater reliabilities: Pearson correlation coefficients and Spearman coefficients (kwo maraco)	Cronbach's alpha: 0.93; Spearman-brown: 0.93; Test-retest (r): 0.84; Inter-rater: 0.74
Betancourt, Yang, Bolton, & Normand (2014)	Multi-step IRT-based structural equation modeling (SEM): examine both the latent structure of the scale and the individual observed items that comprise latent construct; Exploratory factor analysis (EFA); modification indices; confirmatory factor analysis (CFA); correlations between IRT scores (θ) for each mental health problem subscale and ratings of functional impairment; Fisher's Information	IRT analyses: A four factor solution – internalizing problems, externalizing problems, prosocial attitudes/behaviors, and somatic complaints without medical cause – best fit the data; Concurrent validity (with impairment in functioning): internalizing emotional and behavioral problems= 0.36, externalizing problems and somatic complaints without medical cause= 0.28 and subscale for prosocial attitudes and behavior= 0.07, $p = 0.15$; The most total information (area under the subscale information curves): internalizing subscale > externalizing subscale > the prosocial subscale and the subscale for somatic complaints without medical cause; Cronbach's alpha: prosocial/adaptive $\alpha = 0.72$, somatic complaints without medical cause $\alpha = 0.74$, externalizing problems $\alpha = 0.83$ and internalizing problems $\alpha = 0.88$
Bhana, Rathod, Selohilwe, Kathree, & Petersen (2015)	Internal consistency: Cronbach's alpha; Sensitivity; Specificity	Cronbach's alpha: 0.76; Sensitivity: 49%; Specificity: 94%; PPV: 50%; NPV: 94%
Blair et al. (2017)	Bivariate analyses; Fisher's exact test; CFA; model goodness-of-fit; Chi-square; Cronbach's Alpha	Distinctions emerged between cutoff thresholds within the AUDIT; Factor 1: Chi-Square= 63.39, RMSEA= 0.032, Tucker-Lewis NNFI= 0.995, Bentler CFI= 0.998, NFI= 0.998, Cronbach's alpha= 0.85; Factor 2: Chi-Square= 60.41, RMSEA= 0.031, Tucker-Lewis NNFI= 0.995, Bentler CFI= 0.997, NFI= 0.998, Cronbach's alpha= 0.83; Factor 3: Chi-Square= 47.34, RMSEA= 0.028, Tucker-Lewis NNFI= 0.996, Bentler CFI= 0.998, NFI= 0.999, Cronbach's alpha= 0.76
Carney, Myers, & Louw (2016)	Internal consistency: Cronbach's alpha; Test-retest reliability: ICC coefficients, Kappa coefficients	PESQ scale: ICC: 0.82 (95% CI: 0.77–0.86), α: 0.83; CRAFFT scale: ICC: 0.76 (95% CI: 0.66–0.83), α: 0.82; GAIN SS scale:

		ICC: 0.79 (95% CI: 0.73--0.84), α: 0.80
Charak, de Jong, Berckmoes, Ndayisaba, & Reis (2017)	Confirmatory factor analysis (CFA); comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA)	X2 (242, 231) = 451.90, CFI/TLI = 0.94/0.93, RMSEA = 0.06 [90% CI = 0.05–0.07]; Findings supported a five-factor intercorrelated model of the CTQ after dropping one item related to 'molestation'.
Chibanda et al. (2016)	Tool performance against SCID: i) SSQ-14 against any CMD of depression and/or anxiety; ii) SSQ-14 against depression; iii) PHQ-9 against depression; iv) PHQ-2 against depression; v) GAD-7 against anxiety; Sensitivity; Specificity; positive predictive value (PPV); negative predictive value (NPV); Internal reliability: Cronbach's' α	SSQ-14 (vs. CMD): Sensitivity = 84% (95%CI:78–89%), Specificity = 73% (95%CI:63–81%), Cronbach's alpha = 0.74, PPV = 82% (95%CI:75–88%), NPV = 75% (95%CI:66–83%); SSQ-14 (vs. Depression): Sensitivity = 86% (95%CI:80–92%), Specificity = 70% (95%CI:61–79%), PPV = 79% (95%CI:72–85%), NPV = 79% (95%CI:70–87%); PHQ-9 (vs. Depression): Sensitivity = 85% (95%CI:78–90%), Specificity = 69% (95%CI:59–77%), Cronbach's alpha = 0.86, PPV = 78% (95%CI:71–84%), NPV = 77% (95%CI:68–85%); PHQ-2 (vs. Depression): Sensitivity = 91% (95%CI 86–95%), Specificity = 40% (95%CI:31–50%); GAD-7 (vs. Anxiety): Sensitivity = 89% (95%CI:81–94%), Specificity = 73% (95%CI:65–80%), Cronbach's alpha = 0.87, PPV = 69% (95%CI:60–76%), NPV = 91% (95%CI:84–95%)
Chishinga et al. (2011)	Internal consistency: Cronbach's alpha; External reliability: PPV, Sensitivity, Specificity	CES-D: Cronbach's α: 0.84, Sensitivity: 73% and PPV: 76%; AUDIT: Cronbach's α: 0.98, Sensitivity: 60% and PPV: 60%
Cholera et al. (2014)	Sensitivity, specificity, and categorical likelihood ratios	Sensitivity: 78.7% (95% CI: 64.3–89.3); Specificity of 83.4% (95% CI: 79.1–87.2); Likelihood ratios for the commonly used PHQ-9 categories representing mild (5–9), moderate (10–14), moderately severe (15– 19), and severe (20–27) depression were 0.09, 0.50, 3.89, 6.77, and 22.3 respectively; Post-test probability of depression: 10–14 score = 34.2%, 15–20 score = 47.5%, > 20 score = 75%
Chorwe-Sungani & Chipps (2018)	Reliability: Cronbach's alpha; Instrument agreement: McNemar's test	Cronbach's α= 0.86 (SRQ), 0.85 (HSCL-15), 0.80 (EPDS), 0.70 (PRQ), 0.70 (3-item screen); McNemar's test: PRQ and SRQ, EPDS and HSCL-15, HSCL and PRQ, EPDS and SRQ
Denckla et al. (2017)	Exploratory factor analyses (EFA) on half of study and confirmatory factor analysis (CFA) on second half; Internal validity; External validity	4-factor CFA suggested excellent fit (X2 = 13 115.25, df = 496, p < 0.00); Significant associations among the four NOK factors and all eight YSR-DSM oriented scales, specifically depression which ranged from 0.21 to 0.45; Depression: α =

		0.83; Mixed: α = 0.84; Anxiety: α = 0.83; and Somatic: α = 0.77.
Flisher, Sorsdahl, & Lund (2012)	kappa statistic; intraclass correlation	Parents: Anxiety disorders: Kappa= 0.448 (−0.197–0.699), ICC= 0.468 (0.305–0.605); Agoraphobia: Kappa= 0.789 (−0.507–1.541), ICC= 0.793 (0.702–0.858); MDD: Kappa= 0.662 (−0.041–1.283), ICC= 0.664 (0.532–0.765); ADHD: Kappa= 0.559 (−0.118–0.264), ICC= 0.556 (0.417–0.686); ODD: Kappa= 0.662 (−0.158–1.04), ICC= 0.664 (0.537–0.762) ‖ Youths: Anxiety disorders: Kappa= 0.145 (−0.248–0.531), ICC= 0.157 (0.035–0.337); Agoraphobia: Kappa= 0.579 (−0.398–0.759), ICC= 0.595 (0.452–0.708); MDD: Kappa= 0.661 (0.039–1.282), ICC= 0.664 (0.515–0.773); ADHD: Kappa= 0.227 (−0.046–0.499), ICC= 0.245 (0.001–0.462); ODD: Kappa= 0.385 (0.165–0.937), ICC= 0.387 (0.211–0.538)
Francis, Helander, Kapiga, Weiss, & Grosskurth (2015)	Reliability: Sensitivity and specificity	Sensitivity: 0% (among female college students) to 69% (among male casual labourers); Specificity: 52% (male casual labourers) to 85% (female college students)
Geibel et al., (2016)	Reliability: Cronbach's Alpha, test-retest reliability; Criterion validity: sensitivity, specificity, concurrence between the baseline YSR scores and the psychiatric assessments	Cronbach's alpha: 0.385-0.662; Test-retest: 0.25-0.59; Sensitivity (≥ 6.5): 0.750; Specificity (≥6.5): 0.631
Gelaye et al. (2013)	Reliability: Cronbach's Alpha, Rasch analysis; Construct Validity: Factor analysis; Criterion validity: concordance between the score and psychiatrist diagnosis, ROC, Youden Index	Cronbach's alpha: 0.81; Correlations: 0.57-0.75; Test-retest reliability intraclass correlation coefficient: 0.92; Correlations between PHQ-9 items and the construct were all > 0.60; After verification bias adjustment sensitivity: 71.1% (95%CI: 61.2-83.9%); After verification bias adjustment specificity: 76.6% (95%CI: 74.8-79.5%); None of the items misfit the model according to criteria set a priori
Gelaye et al. (2014)	Bayesian modeling approach	Sensitivity: PHQ-9 vs CITI: 79.8 (64.9–90.8) and PHQ-9 vs SCAN: 71.1 (61.2–83.9); Specificity: PHQ-9 vs CITI: 79.1 (74.7–83.7) and PHQ-9 vs SCAN: 76.6 (74.8–79.5); NPV: PHQ-9 vs CITI: 33.8 and PHQ-9 vs SCAN: 23.5; PPV: PHQ-9 vs CITI: 97.4 and PHQ-9 vs SCAN: 94.5
Habtamu et al.	Convergent validity: Pearson’s correlation coefficient (r);	Cronbach’s alpha (sub-scales and overall WHODAS): 0.88-

(2017)	Internal consistency: Cronbach's alpha; Internal responsiveness: paired sample t-test, effect size (ES); Applicability in Ethiopia: CFA; External validity: comparison with Spearman rank correlation	0.98; Cronbach's alpha (caregivers) 0.82-0.99; Pearson's correlation coefficients (service users, baseline): 0.13-0.22; Pearson's correlation coefficients (caregivers, baseline): 0.20 to 0.34; Pearson's correlation coefficients (service users, follow-up): 0.29 to 0.51; Pearson's correlation coefficients (caregivers, follow-up): 0.40 to 0.53
Hanlon et al. (2015)	Convergent validity: Spearman's Rho coefficient, Kruskal-Wallis test of the equality of medians; Internal consistency: Cronbach's Alpha; Sensitivity; Specificity; PPV; NPV; Cohen's Kappa; Youden's Index; Construct validity: exploratory factor analysis with maximum likelihood extraction and varimax rotation; a multivariable logistic regression analysis was carried out with each scale score dichotomised at the optimal cut-off score for detection as the dependant variable	Cronbach's alpha: 0.84 for PHQ-9, 0.90 for SRQ-20, 0.88 for K10 and 0.85 for K6; Sensitivity: 83.3 for PHQ-9, 83.3 for PHQ-2, 83.3 for SRQ-20, 77.8 for K10 and 77.8 for K6; Specificity: 74.7 for PHQ-9, 60.8 for PHQ-2, 74.3 for SRQ-20, 76.7 for K10 and 73.3 for K6; PPV: 17.1 for PHQ-9, 11.7 for PHQ-2, 16.9 for SRQ-20, 17.3 for K10 and 15.4 for K6; NPV: 98.6 for PHQ-9, 98.3 for PHQ-2, 98.6 for SRQ-20, 98.2 for K10 and 98.1 for K6; Youden's Index: 0.58 for PHQ-9, 0.44 for PHQ-2, 0.58 for SRQ-20, 0.55 for K10 and 0.51 for K6; Kappa: 0.21 for PHQ-9, 0.11 for PHQ-2, 0.20 for SRQ-20, 0.21 for K10 and 0.18 for K6; The PHQ-9 and Kessler 10 were unidimensional on EFA, For the SRQ-20, two factors had Eigenvalues greater than 1.0 and Items 16 (worthlessness) and 18 (fatigue) of the SRQ-20 did not load onto any of the factor model
January & Chimbari (2018)	Reliability: Sensitivity, specificity, PPV and NPV; Concordance: Cohen's Kappa statistic; Internal consistency: Cronbach's alpha	No results
Kagee, Bantjes, Saal, & Sefatsa (2019)	Internal consistency: Cronbach's alpha; Receiver operating characteristic (ROC) curve analysis; Area under the curve (AUC)	Cronbach's alpha: 0.95; Sensitivity: 0.76; Specificity: 0.78; Positive predictive value (PPV): 0.24; NPV: 0.97
Kagotho, Patak-Pietrafesa, Ssewamala, & Kirkbride (2018)	Parallel analysis; Confirmatory factor analyses	One-factor 8-item model: females: (x2 (20) = 29.480; RMSEA = .029 [90% CI: .000e.050]; CFI = .98, TLI = .97) and males: (x2 (20) = 44.24; RMSEA = .039 [90% CI: .023e.055]; CFI = .954; TLI = .935)
Kane, Murray, Bass, Johnson, & Bolton (2016)	Internal consistency: Cronbach's alpha; External reliability: Kruskal-Wallis one-way analysis, Area under the curve, Sensitivity, Specificity	A statistically significant difference in mean SSI scores across the levels of the YSR criterion; Cronbach's alpha: 0.98; K-W analysis: 95.4; Discrimination between low and moderate risk use was adequate for all substance (AUC>0.70);

		Discrimination between moderate and high risk use was poor (AUC<0.70) for all substances.
Marquer et al. (2012)	Internal consistency: Cronbach's alpha and Test-retest and interrater reliability: intra-class correlation coefficients (ICC); Kappa statistics: degree of interrater agreement for each question; External validity: comparing the reported PSYCa 3–6 scores against the clinical psychologists' evaluation	Test-retest reliability: 0.81, CI9% [0.68; 0.89]; Interrater reliability: 0.69, CI 95% [0.45; 0.80]; Cronbach's alpha: 0.61, CI 95% [0.54; 0.65]; Sensitivity: 0.77; Specificity: 0.71; The PSYCa 3–6 positive correlations with CGI severity scale: rho = 0.41
Martin, Fincham, & Kagee (2009)	Receiver operating curve (ROC); Areas under the curve (AUC); Sensitivity; Specificity; positive predictive value (PPV); negative predictive value (NPV); CIDI and PDS associations: Chi-square test; Strength of CIDI and PDS associations: Phi and Cramer's V	Sensitivity= 0.68; Specificity= 0.35; Chi-square analysis= 9.22; Phi and Cramer's V test= 0.33; positive predictive value (PPV)= 76.0%; Negative predictive value (NPV)= 56.7%.
Mellins et al. (2017)	Sensitivity, specificity, Positive predictive value, Negative predictive value, kappa statistic; The McNemar test; Construct validity: SF-36 scores	Sensitivity: 73%; Specificity: 81%; PPV: 47%; NPV: 93%; Kappa: 0.44; SF-36 subscales MH: mean difference=−15.11, 95% CI= (−19.87, −10.35)
Mellins et al., 2018	Internal consistency: Cronbach's alpha, Exploratory factor analysis (EFA)	Cronbach's alphas (wave 1, wave 2): total difficulties (0.74, 0.74);
Morojele et al., 2016	Receiver Operating Characteristic (ROC) curve	Males: AUDIT-C: Sensitivity= 0.984; Specificity= 0.138; PPV= 0.716; NPV= 0.800; AUDIT-3: Sensitivity= 0.906; Specificity= 0.310; PPV= 0.744; NPV= 0.600; AUDIT-QF: Sensitivity= 0.969; Specificity= 0.172; PPV= 0.721; NPV= 0.714; AUDIT-PC: Sensitivity= 0.953; Specificity= 0.828; PPV= 0.924; NPV= 0.889; AUDIT-4: Sensitivity= 0.953; Specificity= 0.552; PPV= 0.824; NPV= 0.842; M-FAST: Sensitivity= 0.953; Specificity= 0.586; PPV= 0.836; NPV= 0.850 || Females: AUDIT-C: Sensitivity= 0.979; Specificity= 0.234; PPV= 0.566; NPV= 0.917; AUDIT-3: Sensitivity= 0.813; Specificity= 0.660; PPV= 0.709; NPV= 0.775; AUDIT-QF: Sensitivity= 0.979; Specificity= 0.298; PPV= 0.588; NPV= 0.933; AUDIT-PC: Sensitivity= 0.979; Specificity= 0.766; PPV= 0.810; NPV= 0.973; AUDIT-4: Sensitivity= 0.917; Specificity= 0.851; PPV= 0.863; NPV= 0.909; M-FAST: Sensitivity= 0.958; Specificity= 0.553; PPV= 0.687; NPV= 0.929;

Muris et al. (2006)	Factor analysis; Internal consistency: Cronbach's alpha	Factor analysis (SCARED): five-factor solution, which accounted for 36.32% of the variance; Cronbach's alphas (SCARED): 0.70; Regression analysis (EMBU-C), mothers' rearing behaviors predicted anxiety: overprotection (beta=0.21), rejection (beta=0.16), and anxious rearing (beta=0.16); Regression analysis (EMBU-C), father's rearing behaviors predicted anxiety: anxious rearing (beta=0.23), rejection (beta=0.18), overprotection (beta=0.17) and emotional warmth (beta=−0.11); Statistical analyses did not reveal marked differences in anxiety and parental rearing behaviors between colored and black youth; 2 (cultural groups: colored/black vs. white)×2 (gender) ANCOVA: F= 117.67
Natamba et al. (2014)	Reliability: Cronbach's alpha; Criterion validity: Sensitivity, specificity, PPV	Cronbach's alpha: 0.92; Sp: 78.5%; PPV: 76.5%; Se: 72.7%.
Nolan et al. (2018)	Pearson's correlation coefficient; Fisher's exact test	PHQ-2 and WHO-5: Pearson's correlation coefficient: −0.3289 ($p < 0.0048$) and Fisher's exact test: 0.028; PHQ-9 and WHO-5: Pearson's correlation coefficient: −0.4463 ($p < 0.0001$) and Fisher's exact test: 0.002.
Odenwald et al. (2007)	Reliability: Cronbach's alpha; Convergent validity: Kappa, sensitivity, specificity	Cronbach's $\alpha = 0.86$; Sensitivity: 0.90; Specificity: 0.90; Kappa = 0.69 ($p < .001$)
Ogle, Koen, & Niehaus (2018)	Reliability: Sensitivity and specificity	Depression: Sensitivity: 60% and Specificity: 90.91%; PTSD: Sensitivity: 37% and Specificity: 72.22%; Panic disorder: Sensitivity: 43.75% and Specificity: 81.25%; Agoraphobia: Specificity: 73.61%; Social Phobia: Specificity: 72.22; Generalised Anxiety: Specificity: 74.6
Ojagbemi, Owolabi, Akinyemi, & Ovbiagele (2017)	Criterion validity: correlation of the scores of each items of HRQOLISP-E with the total HADS-D score using the Pearson product moment correlation (r); HRQOLISP-E item scale correlations; coefficient alpha; split-half reliability	HRQOLISP-E items scale correlations: > 0.8 (0.81–0.93) vs. 0.56–0.68 for the HADS-D; Cronbach's alpha = 0.939 vs 0.742 for the HADS-D; Split-half reliability = 0.899 vs 0.739 for HADS-D
Opakunle, Aloba, Akinsulore, Opakunle, & Fatoye (2018)	Reliability: Cronbach's alpha and the items of total scale correlations; Construct validity: correlational analysis, Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV)	Cronbach's alpha: 0.939; Correlations (OCI-R and MINI suicidality module score): r: 0.344, $P < 0.001$; sensitivity: 0.900; specificity: 0.662; PPV: 0.726; NPV: 0.869

Owoso et al. (2014)	Sensitivity; specificity; positive predictive value [PPV]; negative predictive value [NPV]; Fisher Exact Test analyses: to determine significant differences in group response rates for agreement on each mPRIME item; Fisher Exact Test analyses were utilized to determine significant differences in group response rates for agreement on each mPRIME item.	Sensitivity: 40%; Specificity: 65%; For all assessment items on the mPRIME, true positive case scores were not significantly different from those of false positive cases; Across all items, the true positive group generally scored higher for amount of "agree" responses on mPRIME questions while the agreement response rates for false negative individuals were markedly lower than both true positive and false positive groups; There were 5 correlations between mPRIME screen items and corresponding questions on SIPS out of 10 items
Pence et al. (2012)	Sensitivity; Specificity; Likelihood ratio positive (LR+); Likelihood ratio negative (LR−); PPV; NPV	Sensitivity= 27% (95% CI: 6–61%); Specificity= 94% (95% CI: 91–96%) relative to an MDE diagnosis on the CIDI in the past month; LR+= 4.5; LR−= 0.8; PPV= 12% (2–30%); NPV= 98% (96–99%)
Saal, Kagee, & Bantjes (2018)	Reliability: Cronbach's alpha; Receiver operating characteristic (ROC) curve analysis: ability of the BDI-I to discriminate between MDD caseness or non-caseness with the SCID as gold standard; The area under the curve (AUC); to discriminate between those individuals with MDD and those individuals without MDD	Cronbach's alpha: 0.92; Sensitivity: 0.67; Specificity: 0.67; Area under the curve (AUC): 0.77
Seth et al. (2015)	Sensitivity; Specificity	AUDIT-C: Sensitivity= 95.2% (95% CI: 91.9%, 98.6%), Specificity= 77.5% (95% CI: 72.5%, 82.5%); AUDIT-3: Sensitivity= 85.3% (95% CI: 79.7, 91.0), Specificity= 34.1% (95% CI: 28.4%, 39.7%); AUROC: Significantly superior to the line of identity (x-axis = 1—y-axis)
Shanahan, Anderson, & Mkhize (2001)	Internal consistency: Cronbach's alpha; Item-subscale correlations; Test-retest reliability; Scale-level and item-level bilingual technique analysis; Factor analysis	Cronbach's alpha: 0.78; Zero item-subscale correlations; Test-retest coefficients: 0.46 - 0.90; For both versions, all 9 subscales loaded on the first factor which ranged from .77 (Hostility) to .91 (Anxiety) on the English version, and from .72 (Somatization) to .90 (Anxiety) on the Zulu version
Sharp et al. (2014)	Two-dimensional ROC; Sensitivity; 1-Specificity; AUC	Cronbach's alpha: Youth report= 0.26–0.62, Caregiver report= 0.30–0.72, teacher report= 0.67–0.84; Intraclass correlations: Caregiver and self-report= 0.46, caregiver and teacher= 0.30, self-report and teacher= 0.07; ROC: Caregiver= 0.76, Youth= 0.58, Teacher= 0.56; Sensitivity: Caregiver= 0.70, Youth= 0.58, Teacher= 0.58; Specificity: Caregiver= 0.73, Youth=

		0.51, Teacher= 0.56
Stewart et al. (2009)	Sensitivity; Specificity; Positive predictive value (ppv); Negative predictive value (npv); Receiver Operating Characteristic (ROC) curves; Cronbach's alpha	Major Depression: Specificity= 80.8, Sensitivity= 66.2, PPV= 36.1, NPV= 93.6; Major/Minor Depression: Specificity= 87.6, Sensitivity= 56.4, PPV= 66.7, NPV= 82.1; Cronbach's alpha 0.85
Stewart, Umar, Tomenson, & Creed (2013)	Sensitivity; Specificity; PPV; NPV; Cronbach's alpha, ROC curves; AUC	SRQ (major depression): AUC= 0.833 (95% CI 0.770-0.897), Sp= 76.4, Se= 61.8, PPV= 32.7, NPV= 90.7; EPDS (major depression): AUC= 0.811 (95% CI 0.734-0.889), Sp= 81.3, Se= 76.3, PPV= 40.0, NPV= 95.7; SRQ (minor depression): AUC= 0.883 (95% CI 0.839-0.927), Sp= 78.7, Se= 83.5, PPV= 54.4, NPV= 94.3; EPDS (minor depression): AUC= 0.767 (95% CI 0.695-0.839), Sp= 85.8, Se= 62.5, PPV= 60.0, NPV= 87.2
Suliman, Kaminer, Seedat, & Stein (2005)	Cohen's kappa coefficients (K): the level of agreement between the measures	Level of agreement: 0.74 (SE = 0.15; CI = 0.46–1.0).
Tomita, Kandolo, Susser, & Burns (2016)	Reliability: test and retest method (among 135 who participated in both the baseline and follow-up interview); weighted Kappa coefficient; intraclass correlation coefficient; two-way random effect model; descriptive statistics.	Intraclass correlation coefficient two-way random effect model: 0.45; Weighted Kappa: 0.25; Over three-quarters rated the experience as ‘good’ or ‘excellent’, and approximately 95% would recommend it to their family and friends
Udedi, Muula, Stewart, & Pence (2019)	Sensitivity; Specificity; PPV; NPV; Internal Consistency: Cronbach's Alpha	Sensitivity: 64%; Specificity of 94%; NPV: 79%; PPV: 88%
van Heyningen, Honikman, Tomlinson, Field, & Myer (2018)	Internal consistency: Cronbach’s alpha and interrater reliability: pearson's correlation matrix; Sensitivity; Specificity; PPV; NPV;	Cronbach's alpha (K10, K6, PHQ9 and Whooley questions with the help question): >0.8, cronbach's alpha (EPDS and Whoole questions w/out help question): 0.7-08 and cronbach's alpha (EPDS-3, PHQ-2 and anxiety subscale): 0.5-0.6; Pearson’s correlation coefficient: 0.60 (GAD and EPDS) to 0.73 (K10 and EPDS); Sensitivity (MDE): 65%-89%; Specificity (MDE): 74%-87%; Sensitivity (Anxiety): 64%-76%; Specificity (Anxiety): 59%-76%; Sensitivity (MDE & Anxiety): 66%-80%; Specificity (MDE & Anxiety): 76%-85%
Westhuizen,	Internal consistency: Cronbach's alpha; Sensitivity;	Cronbach's alpha: TSI = 0.87, Alcohol and Cannabis = 0.81,

Wyatt, Williams, Stein, & Sorsdahl (2016)	Specificity	Tik = 0.95, Mandrax = 0.90; Mann–Whitney U-test: Alcohol = 505.5, Cannabis = 1876, Tik = 2278.5, Mandrax = 2112.5; TSI (Use/Abuse): Sensitivity = 79%, Specificity = 72%; TSI (Abuse/Dependence): Sensitivity = 64%, Specificity = 61%; SSI-Alcohol (Use/Abuse): Sensitivity = 85%, Specificity = 89%; SSI-Alcohol (Abuse/Dependence): Sensitivity = 60%, Specificity = 60%; SSI-Illicit Drugs (Use/Abuse): Sensitivity = 93%, Specificity = 93%; SSI-Illicit Drugs (Abuse/Dependence): Sensitivity = 90%, Specificity = 87%
Woldetensay et al. (2018)	Reliability: Cronbach's Alpha, test-retest reliability (ICC); Criterion validity: sensitivity, specificity, exploratory factor analysis (EFA), ROC, Rasch analysis	Cronbach's alpha: 0.84; ICC: 0.30-0.54; One-week test-retest reliability: 0.98 for both agreement and consistency ICC indices; The quadratic and linear weighted kappa: 0.97-0.86; Sensitivity (≥8): 80.8%; Specificity (≥8) 79.5%.

VIII. References

- Adewuya, A. O., Ola, B. A., & Afolabi, O. O. (2006). Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. *Journal of Affective Disorders*, *96*(1), 89–93. https://doi.org/10.1016/j.jad.2006.05.021
- Africa Conflict Prevention Pool (Great Britain), & Marshall, M. G. (2005). *Conflict Trends in Africa, 1946-2004: A Macro-comparative Perspective*(Rep.). Center for Systemic Peace.
- Akena, D., Joska, J., Obuku, E. A., & Stein, D. J. (2013). Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care*, *25*(10), 1245–1252. https://doi.org/10.1080/09540121.2013.764385
- Akyeampong, E., Hill, A. G., & Kleinman, A. M. (Eds.). (2015). The culture of mental illness and psychiatric practice in africa. Retrieved from https://ebookcentral.proquest.com
- Ali, G., Ryan, G., & Silva, M. J. (2016). Validated Screening Tools for Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. Plos One,11(6). doi:10.1371/journal.pone.0156939
- American Academy of Pediatrics. (2012). Mental Health Screening and Assessment Tools for Primary Care(Rep.).
- Andersen, L. S., Grimsrud, A., Myer, L., Williams, D. R., Stein, D. J., & Seedat, S. (2011). The psychometric properties of the K10 and K6 scales in screening for mood and

anxiety disorders in the South African Stress and Health study. *International Journal of Methods in Psychiatric Research*, *20*(4), 215–223. https://doi.org/10.1002/mpr.351

- Barthel, D., Barkmann, C., Ehrhardt, S., Schoppen, S., & Bindt, C. (2015). Screening for depression in pregnant women from Côte d′Ivoire and Ghana: Psychometric properties of the Patient Health Questionnaire-9. *Journal of Affective Disorders*, *187*, 232–240. https://doi.org/10.1016/j.jad.2015.06.042
- Berard, R. M. F., Boermeester, F., & Viljoen, G. (1998). Depressive disorders in an out-patient oncology setting: prevalence, assessment, and management. *Psycho-Oncology*, *7*(2), 112–120. https://doi.org/10.1002/(SICI)1099-1611(199803/04)7:2<112::AID-PON300>3.0.CO;2-W
- Bertschy, G., Viel, J. F., & Ahyi, R. G. (1992). Depression in Benin: An assessment using the Comprehensive Psychopathological Rating Scale and the principal component analysis. *Journal of Affective Disorders*, *25*(3), 173–180. https://doi.org/10.1016/0165-0327(92)90003-O
- Betancourt, T. S., Yang, F., Bolton, P., & Normand, S.-L. (2014). Developing an African youth psychosocial assessment: an application of item response theory. *International Journal of Methods in Psychiatric Research*, *23*(2), 142–160. https://doi.org/10.1002/mpr.1420
- Betancourt, T., 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. https://doi.org/10.1007/s00127-008-0475-1

- Bhana, A., Rathod, S. D., Selohilwe, O., Kathree, T., & Petersen, I. (2015). The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. *BMC Psychiatry*, *15*. https://doi.org/10.1186/s12888-015-0503-0
- Blair, A. H., Pearce, M. E., Katamba, A., Malamba, S. S., Muyinda, H., Schechter, M. T., & Spittal, P. M. (2017). The Alcohol Use Disorders Identification Test (AUDIT): Exploring the Factor Structure and Cutoff Thresholds in a Representative Post-Conflict Population in Northern Uganda. *Alcohol and Alcoholism*, *52*(3), 318–327. https://doi.org/10.1093/alcalc/agw090
- Breuer, E., Myer, L., Struthers, H., & Joska, J. A. (2011). HIV/AIDS and mental health research in sub-Saharan Africa: a systematic review. African Journal of AIDS Research (AJAR), 10(2), 101–122. https://doi.org/10.2989/16085906.2011.593373
- Butler, M., Kane, R. L., McAlpine, D., Kathol, R. G., Fu, S. S., Hagedorn, H., & Wilt, T. J. (2008). Integration of Mental Health/Substance Abuse and Primary Care (No. 173). Rockville, MD: Agency for Healthcare Research and Quality (US).
- Carney, T., Myers, B., & Louw, J. (2016). Reliability of the GAIN-SS, CRAFTT and PESQ screening instruments for substance use among South African adolescents. *The South African Journal of Psychiatry : SAJP : The Journal of the Society of Psychiatrists of South Africa*, *22*(1). https://doi.org/10.4102/sajpsychiatry.v22i1.932
- Charak, R., de Jong, J. T. V. M., Berckmoes, L. H., Ndayisaba, H., & Reis, R. (2017). Assessing the factor structure of the Childhood Trauma Questionnaire, and cumulative effect of abuse and neglect on mental health among adolescents in conflict-affected

Burundi. *Child Abuse & Neglect, 72*, 383–392. https://doi.org/10.1016/j.chiabu.2017.09.009

- Chibanda, D., Verhey, R., Gibson, L. J., Munetsi, E., Machando, D., Rusakaniko, S., … Abas, M. (2016). Validation of screening tools for depression and anxiety disorders in a primary care population with high HIV prevalence in Zimbabwe. *Journal of Affective Disorders*, *198*, 50–55. https://doi.org/10.1016/j.jad.2016.03.006
- Chishinga, N., Kinyanda, E., Weiss, H. A., Patel, V., Ayles, H., & Seedat, S. (2011). Validation of brief screening tools for depressive and alcohol use disorders among TB and HIV patients in primary care in Zambia. *BMC Psychiatry*, *11*, 75. https://doi.org/10.1186/1471-244X-11-75
- Cholera, R., Gaynes, B. N., Pence, B. W., Bassett, J., Qangule, N., Macphail, C., … Miller, W. C. (2014). Validity of the patient health questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *Journal of Affective Disorders*, *167*, 160–166. https://doi.org/10.1016/j.jad.2014.06.003
- Chorwe-Sungani, G., & Chipps, J. (2018). Performance of the 3-item screener, the Edinburgh Postnatal Depression Scale, the Hopkins Symptoms Checklist-15 and the Self-Reporting Questionnaire and Pregnancy Risk Questionnaire, in screening of depression in antenatal clinics in the Blantyre district of Malawi. *Malawi Medical Journal*, *30*(3), 184–190. https://doi.org/10.4314/mmj.v30i3.10
- Denckla, C. A., Ndetei, D. M., Mutiso, V. N., Musyimi, C. W., Musau, A. M., Nandoya, E. S., … McKenzie, K. (2017). Psychometric properties of the Ndetei–Othieno–Kathuku (NOK) Scale: A mental health assessment tool for an African setting. *Journal of Child*

and Adolescent Mental Health, *29*(1), 39–49. https://doi.org/10.2989/17280583.2017.1310729

- Egger, M., & Smith, G. D. (1998). Meta-analysis bias in location and selection of studies. Bmj,316(7124), 61-66. doi:10.1136/bmj.316.7124.61
- Flisher, A. J., Sorsdahl, K. R., & Lund, C. (2012). Test–retest reliability of the Xhosa version of the Diagnostic Interview Schedule for Children. *Child: Care, Health and Development*, *38*(2), 261–265. https://doi.org/10.1111/j.1365-2214.2010.01195.x
- Francis, J. M., Helander, A., Kapiga, S. H., Weiss, H. A., & Grosskurth, H. (2015). Validation of the MINI (DSM IV) Tool for the Assessment of Alcohol Dependence among Young People in Northern Tanzania Using the Alcohol Biomarker Phosphatidylethanol (PEth). *International Journal of Environmental Research and Public Health*, *12*(11), 14021–14033. https://doi.org/10.3390/ijerph121114021
- Funk, M., Saraceno, B., Pathare, S., Flisher, A. J., & World Health Organization (Eds.). (2003). The Mental Health Context. Geneva: World Health Organization.
- Geibel, S., Habtamu, K., Mekonnen, G., Jani, N., Kay, L., Shibru, J., … Kalibala, S. (2016). Reliability and Validity of an Interviewer-Administered Adaptation of the Youth Self-Report for Mental Health Screening of Vulnerable Young People in Ethiopia. *PLoS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0147267
- Gelaye, B., Tadesse, M. G., Williams, M. A., Fann, J. R., Stoep, A. V., & Zhou, X.-H. A. (2014). Assessing Validity of a Depression Screening Instrument in the Absence of a Gold Standard. *Annals of Epidemiology*, *24*(7), 527–531. https://doi.org/10.1016/j.annepidem.2014.04.009

- Gelaye, B., Williams, M. A., Lemma, S., Deyessa, N., Bahretibeb, Y., Shibre, T., … Zhou, X.-H. A. (2013). Validity of the Patient Health Questionnaire-9 for Depression Screening and Diagnosis in East Africa. *Psychiatry Research*, *210*(2). https://doi.org/10.1016/j.psychres.2013.07.015
- Gender and women's mental health. (2013, June 24). Retrieved from https://www.who.int/mental_health/prevention/genderwomen/en/
- Global HIV and AIDS statistics. (2018, October 08). Retrieved from https://www.avert.org/global-hiv-and-aids-statistics
- Habtamu, K., Alem, A., Medhin, G., Fekadu, A., Dewey, M., Prince, M., & Hanlon, C. (2017). Validation of the World Health Organization Disability Assessment Schedule in people with severe mental disorders in rural Ethiopia. *Health and Quality of Life Outcomes*, *15*. https://doi.org/10.1186/s12955-017-0647-3
- Hanlon, C., Medhin, G., Selamu, M., Breuer, E., Worku, B., Hailemariam, M., … Fekadu, A. (2015). Validity of brief screening questionnaires to detect depression in primary care in Ethiopia. *Journal of Affective Disorders*, *186*, 32–39. https://doi.org/10.1016/j.jad.2015.07.015
- January, J., & Chimbari, M. J. (2018). Study protocol on criterion validation of Edinburgh Postnatal Depression Scale (EPDS), Patient Health Questionnaire (PHQ-9) and Centre for Epidemiological Studies-Depression (CES-D) screening tools among rural postnatal women; a cross-sectional study. *BMJ Open*, *8*(4), e019085. https://doi.org/10.1136/bmjopen-2017-019085
- Johnson, S. (2016, April 26). Africa's life expectancy jumps dramatically. Retrieved from https://www.ft.com/content/38c2ad3e-0874-11e6-b6d3-746f8e9cdd33

- Kagee, A., Bantjes, J., Saal, W., & Sefatsa, M. (2019). Utility of the Posttraumatic Stress Scale–Self-report version in screening for posttraumatic stress disorder among persons seeking HIV testing. *South African Journal of Psychology*, *49*(1), 136–147. https://doi.org/10.1177/0081246318779191
- Kagotho, N., Patak-Pietrafesa, M., Ssewamala, F. M., & Kirkbride, G. (2018). Assessing the Association Between Depression and Savings for Kenyan Youth Using a Validated Child Depression Inventory Measure. *Journal of Adolescent Health*, *62*(1, Supplement), S21–S28. https://doi.org/10.1016/j.jadohealth.2017.11.002
- Kane, J. C., Murray, L. K., Bass, J. K., Johnson, R. M., & Bolton, P. (2016). Validation of a substance and alcohol use assessment instrument among orphans and vulnerable children in Zambia using Audio Computer Assisted Self-Interviewing (ACASI). *Drug and Alcohol Dependence*, *166*, 85–92. https://doi.org/10.1016/j.drugalcdep.2016.06.026
- Marquer, C., Barry, C., Mouchenik, Y., Hustache, S., Djibo, D. M., Manzo, M. L., … Moro, M.-R. (2012). A rapid screening tool for psychological distress in children 3–6years old: results of a validation study. *BMC Psychiatry*, *12*, 170. https://doi.org/10.1186/1471-244X-12-170
- Martin, L., Fincham, D., & Kagee, A. (2009). Screening for HIV-related PTSD: Sensitivity and specificity of the 17-item Posttraumatic Stress Diagnostic Scale (PDS) in identifying HIV-related PTSD among a South African sample. *African Journal of Psychiatry*, *12*, 270–274. https://doi.org/10.4314/ajpsy.v12i4.49041
- Mellins, Claude A., Kauchali, S., Nestadt, D. F., Bai, D., Aidala, A., Myeza, N., … Davidson, L. L. (2017). Validation of the Client Diagnostic Questionnaire to Assess

Mental Health in South African Caregivers of Children. *Clinical Psychology & Psychotherapy, 24*(1), 245–254. https://doi.org/10.1002/cpp.2008

- Mellins, Claude Ann, Xu, Q., Nestadt, D. F., Knox, J., Kauchali, S., Arpadi, S., … Davidson, L. L. (2018). Screening for Mental Health Among Young South African Children: the Use of the Strengths and Difficulties Questionnaire (SDQ). *Global Social Welfare*, *5*(1), 29–38. https://doi.org/10.1007/s40609-018-0110-3
- Miller, J. E. (2014). The Need for Early Mental Health Screening and Intervention Across the Lifespan (Issue brief). Alexandria, VA: American Mental Health Counselors Association.
- Mojtabai, R., Olfson, M., Sampson, N. A., Jin, R., Druss, B., Wang, P. S., … Kessler, R. C. (2011). Barriers to Mental Health Treatment: Results from the National Comorbidity Survey Replication (NCS-R). Psychological Medicine, 41(8), 1751–1761. https://doi.org/10.1017/S0033291710002291
- Morojele, N. K., Nkosi, S., Kekwaletswe, C. T., Shuper, P. A., Manda, S. O., Myers, B., & Parry, C. D. H. (2016). Utility of Brief Versions of the Alcohol Use Disorders Identification Test (AUDIT) to Identify Excessive Drinking Among Patients in HIV Care in South Africa. *Journal of Studies on Alcohol and Drugs*, *78*(1), 88–96. https://doi.org/10.15288/jsad.2017.78.88
- Muris, P., Loxton, H., Neumann, A., du Plessis, M., King, N., & Ollendick, T. (2006). DSM-defined anxiety disorders symptoms in South African youths: Their assessment and relationship with perceived parental rearing behaviors. *Behaviour Research and Therapy*, *44*(6), 883–896. https://doi.org/10.1016/j.brat.2005.06.002

- Mutumba, M., Tomlinson, M., & Tsai, A. C. (2014). Psychometric properties of instruments for assessing depression among African youth: A systematic review. Journal of Child & Adolescent Mental Health,26(2), 139-156. doi:10.2989/17280583.2014.907169
- NAMI. (n.d.). Mental Health Screening. Retrieved from https://www.nami.org/learn-more/public-policy/mental-health-screening
- Natamba, B. K., Achan, J., Arbach, A., Oyok, T. O., Ghosh, S., Mehta, S., … Young, S. L. (2014). Reliability and validity of the center for epidemiologic studies-depression scale in screening for depression among HIV-infected and -uninfected pregnant women attending antenatal services in northern Uganda: a cross-sectional study. *BMC Psychiatry, 14*. https://doi.org/10.1186/s12888-014-0303-y
- Nolan, C. P., O'Donnell, P. J. M., Desderius, B. M., Mzombwe, M., McNairy, M. L., Peck, R. N., & Kingery, J. R. (2018). Depression screening in HIV-positive Tanzanian adults: comparing the PHQ-2, PHQ-9 and WHO-5 questionnaires. *Global Mental Health, 5*. https://doi.org/10.1017/gmh.2018.31
- O'Connor, E., Rossom, R. C., Henninger, M., Groom, H. C., & Burda, B. U. (2016). Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA, 315(4), 388–406. https://doi.org/10.1001/jama.2015.18948
- Odenwald, M., Lingenfelder, B., Schauer, M., Neuner, F., Rockstroh, B., Hinkel, H., & Elbert, T. (2007). Screening for Posttraumatic Stress Disorder among Somali ex-combatants: A validation study. *Conflict and Health, 1*, 10. https://doi.org/10.1186/1752-1505-1-10

- Ogle, Z., Koen, L., & Niehaus, D. J. H. (2018). The validation of the visual screening tool for anxiety disorders and depression in hypertension and/or diabetes. *African Journal of Primary Health Care & Family Medicine*, *10*(1), 7. https://doi.org/10.4102/phcfm.v10i1.1722
- Ojagbemi, A., Owolabi, M., Akinyemi, J., & Ovbiagele, B. (2017). Proposing a new stroke-specific screening tool for depression: Examination of construct validity and reliability. *ENeurologicalSci*, *9*, 14–18. https://doi.org/10.1016/j.ensci.2017.10.002
- Opakunle, T., Aloba, O., Akinsulore, A., Opakunle, O., & Fatoye, F. (2018). Obsessive-compulsive inventory-revised: Factor structure, reliability, validity, and suicide risk screening characteristics among nigerian patients with schizophrenia. *Journal of Neurosciences in Rural Practice*, *9*(2), 219. https://doi.org/10.4103/jnrp.jnrp_538_17
- Owoso, A., Ndetei, D. M., Mbwayo, A. W., Mutiso, V. N., Khasakhala, L. I., & Mamah, D. (2014). Validation of a modified version of the PRIME screen for psychosis-risk symptoms in a non-clinical Kenyan youth sample. *Comprehensive Psychiatry*, *55*(2), 380–387. https://doi.org/10.1016/j.comppsych.2013.10.004
- Pence, B. W., Gaynes, B. N., Atashili, J., O'Donnell, J. K., Tayong, G., Kats, D., … Ndumbe, P. M. (2012). Validity of an interviewer-administered patient health questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *Journal of Affective Disorders*, *143*(1), 208–213. https://doi.org/10.1016/j.jad.2012.05.056
- Saal, W., Kagee, A., & Bantjes, J. (2018). Utility of the Beck Depression Inventory in measuring major depression among individuals seeking HIV testing in the Western Cape, South Africa. *AIDS Care*, *30*(sup1), 29–36. https://doi.org/10.1080/09540121.2018.1499856

- Seth, P., Glenshaw, M., Sabatier, J. H. F., Adams, R., Du Preez, V., DeLuca, N., & Bock, N. (2015). AUDIT, AUDIT-C, and AUDIT-3: drinking patterns and screening for harmful, hazardous and dependent drinking in Katutura, Namibia. *PloS One*, *10*(3), e0120850. https://doi.org/10.1371/journal.pone.0120850
- Shanahan, S. F., Anderson, S. J., & Mkhize, N. J. (2001). Assessing Psychological Distress in Zulu-Speakers: Preliminary Findings from an Adaptation of the SCL-90-R. *South African Journal of Psychology*, *31*(4), 1–12. https://doi.org/10.1177/008124630103100401
- Sharp, C., Venta, A., Marais, L., Skinner, D., Lenka, M., & Serekoane, J. (2014). First Evaluation of a Population-Based Screen to Detect Emotional-Behavior Disorders in Orphaned Children in Sub-Saharan Africa. *AIDS and Behavior*, *18*(6), 1174–1185. https://doi.org/10.1007/s10461-014-0739-6
- Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., … Pignone, M. P. (2016). Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. JAMA, 315(4), 380–387. https://doi.org/10.1001/jama.2015.18392
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. Journal of Research in Science Teaching,38(5), 553-573. doi:10.1002/tea.1018
- Stewart, R. C., Kauye, F., Umar, E., Vokhiwa, M., Bunn, J., Fitzgerald, M., … Creed, F. (2009). Validation of a Chichewa version of the Self-Reporting Questionnaire (SRQ) as a brief screening measure for maternal depressive disorder in Malawi, Africa. *Journal of Affective Disorders*, *112*(1), 126–134. https://doi.org/10.1016/j.jad.2008.04.001

- Stewart, R. C., Umar, E., Tomenson, B., & Creed, F. (2013). Validation of screening tools for antenatal depression in Malawi—A comparison of the Edinburgh Postnatal Depression Scale and Self Reporting Questionnaire. *Journal of Affective Disorders*, *150*(3), 1041–1047. https://doi.org/10.1016/j.jad.2013.05.036
- Stiffler, M. C., & Dever, B. V. (2015). History of Screening Practices, Mental Health Assessment, and Classification in the USA. In M. C. Stiffler & B. V. Dever (Eds.), Mental Health Screening at School: Instrumentation, Implementation, and Critical Issues (pp. 5–26). https://doi.org/10.1007/978-3-319-19171-3_2
- Stigma and discrimination. (2015, August 7). Retrieved April 21, 2019, from Mental Health Foundation website: https://www.mentalhealth.org.uk/a-to-z/s/stigma-and-discrimination
- Suliman, S., Kaminer, D., Seedat, S., & Stein, D. (2005). Assessing post-traumatic stress disorder in South African adolescents: using the child and adolescent trauma survey (CATS) as a screening tool. *Annals of General Psychiatry*, *4*(1), 2. https://doi.org/10.1186/1744-859X-4-2
- Sweetland, A. C., Belkin, G. S., & Verdeli, H. (2013). Measuring Depression And Anxiety In Sub-Saharan Africa. Depression and Anxiety,31(3), 223-232. doi:10.1002/da.22142
- The World Bank. (n.d.). No poverty. Retrieved from http://datatopics.worldbank.org/sdgatlas/archive/2017/SDG-01-no-poverty.html
- The World Bank. (n.d.). World Bank Country and Lending Groups. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

- Tomita, A., Kandolo, K. M., Susser, E., & Burns, J. K. (2016). Use of short messaging services to assess depressive symptoms among refugees in South Africa: Implications for social services providing mental health care in resource-poor settings. *Journal of Telemedicine and Telecare*, *22*(6), 369–377. https://doi.org/10.1177/1357633X15605406
- Tsai, A. C. (2014). Reliability and Validity of Depression Assessment Among Persons With HIV in Sub-Saharan Africa. JAIDS Journal of Acquired Immune Deficiency Syndromes,66(5), 503-511. doi:10.1097/qai.0000000000000210
- Tsai, A. C., Scott, J. A., Hung, K. J., Zhu, J. Q., Matthews, L. T., Psaros, C., & Tomlinson, M. (2013). Reliability and Validity of Instruments for Assessing Perinatal Depression in African Settings: Systematic Review and Meta-Analysis. PLoS ONE,8(12). doi:10.1371/journal.pone.0082521
- U.S National Library of Medicine. (2019, February 15). Mental Disorders. Retrieved from https://medlineplus.gov/mentaldisorders.html
- U.S. National Library of Medicine. (2019, January 31). Mental Health Screening: MedlinePlus Lab Test Information. Retrieved from https://medlineplus.gov/lab-tests/mental-health-screening/
- Udedi, M., Muula, A. S., Stewart, R. C., & Pence, B. W. (2019). The validity of the patient health Questionnaire-9 to screen for depression in patients with type-2 diabetes mellitus in non-communicable diseases clinics in Malawi. *BMC Psychiatry*, *19*(1), 81. https://doi.org/10.1186/s12888-019-2062-2
- van Heyningen, T., Honikman, S., Tomlinson, M., Field, S., & Myer, L. (2018). Comparison of mental health screening tools for detecting antenatal depression and

anxiety disorders in South African women. *PLoS ONE, 13*(4). https://doi.org/10.1371/journal.pone.0193697

- Westhuizen, C. van der, Wyatt, G., Williams, J. K., Stein, D. J., & Sorsdahl, K. (2016). Validation of the Alcohol, Smoking and Substance Involvement Screening Test in a low- and middle-income country cross-sectional emergency centre study. *Drug and Alcohol Review*, *35*(6), 702–709. https://doi.org/10.1111/dar.12424
- Wilson, J. M. G., Jungner, G., & Organization, W. H. (1968). Principles and practice of screening for disease. Retrieved from https://apps.who.int/iris/handle/10665/37650
- Woldetensay, Y. K., Belachew, T., Tesfaye, M., Spielman, K., Biesalski, H. K., Kantelhardt, E. J., & Scherbaum, V. (2018). Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo version. *PLoS ONE*, *13*(2). https://doi.org/10.1371/journal.pone.0191782
- World Health Organization, & UNAIDS. (2001). World Health Report 2001 Mental Health: New Understanding, New Hope. Albany: World Health Organization.
- World Health Organization. (2004). ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (2nd ed). Geneva: World Health Organization. http://www.who.int/iris/handle/10665/42980
- World Health Organization. (2010, November 05). Process of translation and adaptation of instruments. Retrieved from https://www.who.int/substance_abuse/research_tools/translation/en/
- World Health Organization. (2014, August 15). Mental health: A state of well-being. Retrieved from https://www.who.int/features/factfiles/mental_health/en/

- World Health Organization. (2017). Depression and other common mental disorders: global health estimates. World Health Organization. http://www.who.int/iris/handle/10665/254610. License: CC BY-NC-SA 3.0 IGO
- World Health Organization. (2018, April 9). Mental disorders. Retrieved from https://www.who.int/news-room/fact-sheets/detail/mental-disorders