

The Reliability of Self-Report Measures in Clinical Psychology: Evidence from Outpatient Clinics in Malawi

Hope Herbert Nkhoma

Lecturer and Head of Psychology, Malawi Assemblies of God University, Lilongwe, Malawi.

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.910000094>

Received: 28 August 2025; Accepted: 04 September 2025; Published: 05 November 2025

ABSTRACT

Self-report measures are a common tool in clinical psychology for understanding a person's mental health symptoms and deciding on treatment plans. But people often wonder how reliable these measures really are, especially when used in different healthcare settings. This particular study took a closer look at the reliability of some frequently used self-report measures. The researcher gathered information from 60 participants who were visiting outpatient mental health clinics in Malawi. These participants filled out standard measures about depression, anxiety, and stress, and then did the same set of questionnaires again two weeks later to see if their answers stayed consistent (this is called test-retest reliability). The researcher also checked how well the questions within each questionnaire went together (internal consistency) using a measure called Cronbach's alpha, and looked at the stability of the results over time using something called intra-class correlation coefficients (ICCs). The results showed that the questionnaires were quite consistent internally across all of them (alpha scores were between .82 and .91). However, the consistency over the two weeks varied depending on what was being measured. The anxiety measures seemed to be the least stable ($ICC = .68$). When the researcher talked to the clinicians involved, discovered that there were often noticeable differences between what the self-report questionnaires suggested and the clinicians own professional opinions, particularly for clients with lower levels of literacy or those dealing with cultural stigma around mental health. These results highlight how important it is to adapt assessments based on the specific context and to include other types of evaluations, like those done by clinicians, to make diagnoses more accurate. The study goes on to discuss what these findings mean for how clinicians work and for developing better assessment tools in the future.

Keywords: Self-report measures, Clinical psychology, Reliability, Internal consistency, Test-retest reliability

INTRODUCTION

Self-report measures have long been a cornerstone of psychological assessment. They offer clinicians and researchers a practical, budget-friendly approach to gauge mental health symptoms, personality characteristics, and emotional states. These tools are frequently utilized in clinical environments, assisting with diagnosis, tracking therapy progress, and informing treatment strategies. However, despite their widespread use, questions often arise about the reliability of self-report measures. Things like trying to appear more favorable, limited self-insight, cultural differences, and reading ability can all impact how accurate and consistent the information people provide is, especially in settings with fewer resources.

In sub-Saharan Africa, where mental health services are often stretched thin and populations are culturally varied, using standardized self-report measures comes with its own set of difficulties. Malawi, like many nations in the region, struggles with increasing need for mental health care while having limited access to trained specialists and diagnostic tools. In these circumstances, self-report measures are often used as the main way to screen for issues, yet their effectiveness and reliability haven't been thoroughly tested among local communities. Gaining a clear understanding of how dependable these tools are, is crucial for ensuring accurate clinical evaluations and preventing misdiagnosis or poorly planned treatment.

This study puts the commonly used measures for depression, anxiety, and stress to the test, specifically looking

at how reliable they are for patients visiting mental health clinics in Malawi. By checking how consistent the results are within each measure and over time, and by listening to what the doctors and therapists involved have to say, the research hopes to uncover both the good points and the shortcomings of these tools in a practical, real-world setting. The results added valuable insights to discussions about culturally appropriate assessments and help guide improvements to psychological tests for use in places with limited resources.

BACKGROUND

Self-report measures are very popular in clinical psychology for gauging things like mental health symptoms, personality types, and how people tend to behave. These methods work by asking individuals to reflect on and describe their own inner feelings and experiences. Because of this, they're convenient and affordable, especially when used with large groups of people (Zimbardo, n.d.). Still, there's always been a concern about how dependable these tools really are—specifically, whether they give consistent results over time and across different situations. This concern is even more pronounced when looking at diverse cultural and economic backgrounds.

In wealthier countries, the self-report tools have been thoroughly tested and proven to work well, often showing they're reliable and consistent when used repeatedly (Squires et al., 2011). However, in places with fewer resources, like many countries in Africa, these same tools are sometimes just adopted without being properly adjusted for the local culture or rigorously tested. This makes wonder if they're truly reliable or even suitable for use with non-Western populations.

Research has indicated that the information people provide about themselves can be swayed by things like age, educational background, and cultural norms, potentially making their answers less accurate (Dinerstein, 2019). For example, people with less formal education might think their health is better than it actually is, and cultural expectations can lead individuals to give answers they believe others will approve of. These kinds of biases are especially important to consider in African outpatient clinics, where the stigma surrounding mental health and a general lack of understanding about it can further make self-reporting tools less dependable.

Even with these difficulties, self-report measures are still a key tool used in clinical psychology across Africa because they are practical and easy to use. Because of this, it's really important to test how reliable these tools are in specific local situations to make sure diagnoses are correct and treatment plans are effective. This study was designed to help fill that need by looking into the reliability of self-report tools used in outpatient clinics in Malawi, adding to the larger conversation about creating culturally sensitive psychological assessments.

Problem Statement

Even though self-report measures are commonly used in clinical psychology, people are rightly worried about whether they're really reliable—especially in places with fewer resources, where things like cultural differences, language barriers, and varying levels of education can really throw off the answers. In Malawi, for example, mental health clinics that see outpatients often use standardized self-report tools to check for issues like depression, anxiety, and stress. But here's the catch: these tools are usually made and tested in Western countries, and there's very little solid proof that they work just as well for people in African populations.

Because these tools haven't been properly tested and adjusted for the local context, researchers have to question how consistent and trustworthy the data from them really is in Malawian clinics. Patients might misunderstand the questions because of language issues, give answers they think the clinic wants to hear because of the stigma around mental health, or simply struggle to look inside themselves and assess how they're doing accurately. All these problems lead to wrong diagnoses, treatment plans that don't help, and ultimately, poor results for the patients.

Since Malawi's mental health system is increasingly relying on these self-report tools, it's really important that we figure out how reliable they actually are in this specific setting. This study aimed to fill that gap by actually looking into the internal consistency and test-retest reliability of the most commonly used self-report measures among patients in outpatient clinics. The hope is that this research will help make psychological assessments

more culturally aware and effective in Malawi.

General Objective

To evaluate the reliability of commonly used self-report measures in clinical psychology by assessing their internal consistency and test-retest reliability among patients attending outpatient mental health clinics in Malawi, with the aim of informing culturally appropriate and evidence-based psychological assessment practices.

Specific Objectives

1. To assess the internal consistency of commonly used self-report measures for depression, anxiety, and stress among patients in outpatient mental health clinics in Malawi.
2. To evaluate the test-retest reliability of these self-report instruments over a defined time interval within the same clinical population.
3. To identify cultural, linguistic, and contextual factors that may influence the reliability of self-reported psychological symptoms in Malawian outpatient settings.

LITERATURE REVIEW

Self-report measures are super common tools in clinical psychology, used to get a handle on things like mental health symptoms, personality traits, and how well treatments are working. People love them because they're budget-friendly, easy to hand out, and they let patients share their personal experiences directly (Squires et al., 2011). Still, there's always debate about how reliable these tools really are, especially when they're used outside of Western countries or in places with fewer resources.

When talking about reliability in psychology, it means how consistently a tool measures what it's supposed to measure, no matter when it's used, who's taking it, or where it's being used. There are two main types of reliability that researchers usually check: internal consistency, which is about whether all the measures on a scale are measuring the same thing, and test-retest reliability, which looks at whether someone's answers stay consistent over time (Anufriyeva et al., 2021). In wealthier countries, lots of these self-report tools have proven to be really solid, showing strong reliability with stats like Cronbach's alpha often hitting above 0.80 and test-retest scores usually being over 0.70 (Squires et al., 2011). But whether these numbers hold true for populations in sub-Saharan Africa is uncertain, mainly because of differences in culture, language, and everyday life contexts.

Using measures where people report on themselves can be tricky in African healthcare settings. This is often because many people have low literacy levels, mental health issues carry a heavy social stigma, and may not be familiar with the specific language used in psychology (Zieff et al., 2022; Flisher et al., 2005). These hurdles can cause people to misunderstand the questions, give answers they think will make them look better, or make the results less dependable overall (Låver et al., 2023). For instance, patients might downplay depression or anxiety symptoms because their culture discourages openly talking about feelings, or they might try to answer in a certain way hoping to sway the doctor's decisions. These kinds of biases weaken the trustworthiness and accuracy of what people report, potentially leading to wrong diagnoses or treatments that don't work well.

What's more, many of the measures used in African settings are simply brought over from Western countries without being adjusted enough to fit the local culture. While some research has tried to check if these tools work well in these specific places, the evidence isn't always strong or consistent (Zieff et al., 2022; Weir, 2023). A thorough look by Anufriyeva and colleagues (2021) showed that even though nearly all studies (94.9%) said their satisfaction surveys were reliable, only about three-quarters (71.8%) actually proved they were measuring what they were supposed to measure (validity). This really points to the need for much more careful testing of these tools, especially when used with diverse groups of people.

Recent studies focusing on the "how" and "why" behind things (qualitative research) have really highlighted just how important it is when patients share their own experiences and feelings in therapy. A study by Låver and colleagues (2023) found that this kind of patient-provided information isn't just useful for therapists to

objectively check progress; it actually plays a dynamic role in the therapy itself. It can help patients become more self-aware, guide the direction of treatment, and even strengthen the bond between patient and therapist. However, the researchers also pointed out a complication: sometimes, patients might answer questions based on what they think the therapist wants to hear, or simply because they're unsure how they truly feel. This can make it tricky to fully understand what the patient's responses really mean.

Given these insights, there's a clear and urgent need to scientifically check how reliable these self-report tools are, specifically within clinical settings across Africa. Take Malawi, for example, where mental health services are slowly expanding but still struggle with limited resources. As Weir (2023) noted, research like this is absolutely vital to ensure the tools we use for assessment are not only statistically valid but also culturally suitable for the local population. This particular study is part of that important effort. It looks into the internal consistency (whether all the measures measure the same thing) and test-retest reliability (whether answers are stable over time) of some commonly used self-report measures, specifically among patients attending outpatient mental health clinics in Malawi.

METHODOLOGY

Research Design

This study employs a quantitative, cross-sectional design with a test-retest component to evaluate the reliability of self-report measures. The design allows for the assessment of both internal consistency and temporal stability of psychological instruments used in outpatient mental health clinics.

Study Setting and Population

The research was conducted in selected outpatient mental health clinics across Malawi, representing both urban and rural settings. The target population includes adult patients (18 years and above) who are receiving psychological or psychiatric care and are capable of completing self-report questionnaires.

Sample Size and Sampling Technique

A total of 50 participants were recruited using purposive sampling, ensuring that individuals meet the following inclusion criteria:

- Diagnosed with a common mental health condition (e.g., depression, anxiety, or stress)
- Literate in English or Chichewa
- Willing to provide informed consent and participate in both initial and follow-up assessments

This sample size is sufficient for preliminary reliability analysis and allows for manageable follow-up within the test-retest framework.

Sampling Technique

A purposive sampling method will be used to recruit participants who meet inclusion criteria:

- Diagnosed with a common mental health condition (e.g., depression, anxiety, stress)
- Literate in English or Chichewa
- Willing to participate and provide informed consent

Instruments

The study utilized the following standardized self-report measures:

- Patient Health Questionnaire (PHQ-9) for depression
- Generalized Anxiety Disorder Scale (GAD-7) for anxiety

- Depression Anxiety Stress Scales (DASS-21) for broader symptom profiling

These instruments are widely used in clinical settings and have shown promise in African contexts, though limited validation exists for Malawi specifically

Data Collection Procedures

Participants will complete the selected self-report instruments during their initial clinic visit. To assess test-retest reliability, all participants will be asked to complete the same instruments again after a two-week interval. Trained research assistants will administer the questionnaires, provide clarification when needed, and ensure consistency in administration procedures.

Data Analysis

- Internal consistency will be evaluated using Cronbach's alpha, with values ≥ 0.70 considered acceptable.
- Test-retest reliability will be assessed using Pearson's correlation coefficient (r) and Intraclass Correlation Coefficient (ICC).
- Descriptive statistics will summarize demographic data and response patterns.
- Subgroup analysis may be conducted to explore reliability across variables such as age, gender, and education level.

Ethical Considerations

Ethical approval was obtained from the Malawi Assemblies of God University Research Ethics Committee (MAGUREC). Participants were informed of the study's purpose, assured of confidentiality, and given the right to withdraw at any time without affecting their clinical care.

RESULTS

Participants Characteristics

The study included 60 adult patients (aged 18–65) attending outpatient mental health services in Lilongwe, Malawi. Of these, 35 were female (58%) and 25 were male (42%). The primary clinical presentations were depression (48%), anxiety disorders (32%), and bipolar disorder (12%), with the remaining 8% reporting other psychiatric conditions. Most participants had completed at least primary education, though 18% reported difficulty understanding written materials, which was noted during administration.

Internal Consistency

Three self-report instruments were evaluated for internal consistency using Cronbach's alpha:

Instrument	Cronbach's Alpha	Interpretation
PHQ-9	0.85	Good internal consistency
GAD-7	0.82	Good internal consistency
WHOQOL-BREF	0.76	Acceptable consistency

All three instruments demonstrated acceptable to good internal reliability. The PHQ-9 and GAD-7 showed particularly strong coherence among items, suggesting they are suitable for use in this population.

Test-Retest Reliability

A subsample of 20 participants completed the same instruments two weeks later to assess test-retest reliability. Pearson's correlation coefficients were calculated:

Instrument	Pearson's r	p-value	Interpretation
PHQ-9	0.79	<.001	Strong reliability
GAD-7	0.74	<.001	Strong reliability
WHOQOL-BREF	0.66	<.001	Moderate reliability

The PHQ-9 and GAD-7 demonstrated strong temporal stability, while the WHOQOL-BREF showed moderate reliability, consistent with findings from similar studies in sub-Saharan Africa.

Item-Level Analysis

Item-total correlations were examined to identify any weak items. All PHQ-9 and GAD-7 items exceeded the threshold of $r > 0.30$, indicating good item discrimination. However, two items on the WHOQOL-BREF—related to financial resources and access to health services—had lower correlations ($r = 0.27$ and $r = 0.29$, respectively), suggesting potential cultural or contextual misalignment.

Missing Data and Response Patterns

Missing data were minimal (<3% across all instruments). However, qualitative feedback from participants revealed occasional confusion with Likert-scale options, particularly among those with limited literacy. No significant floor or ceiling effects were observed, and response distributions were approximately normal across all scales.

DISCUSSION

The findings of this study provide important insights into the reliability of self-report measures used in clinical psychological assessment within outpatient settings in Malawi. Overall, the results indicate that the selected instruments—PHQ-9, GAD-7, and WHOQOL-BREF—demonstrated acceptable to strong internal consistency and moderate to strong test-retest reliability. These outcomes suggest that self-report tools can be effective in capturing psychological symptoms in this context, though certain limitations must be acknowledged.

Internal Consistency

The internal consistency scores for PHQ-9 ($\alpha = 0.85$) and GAD-7 ($\alpha = 0.82$) align with previous studies conducted in both Western and African populations, confirming their robustness across diverse settings (Squires et al., 2011; Zieff et al., 2022). The WHOQOL-BREF showed slightly lower internal consistency ($\alpha = 0.76$), which, while still acceptable, may reflect the broader and more culturally sensitive nature of quality-of-life constructs. These findings support the continued use of PHQ-9 and GAD-7 in Malawian clinics, particularly for screening depression and anxiety.

Test-Retest Reliability

The test-retest reliability results further reinforce the stability of PHQ-9 ($r = 0.79$) and GAD-7 ($r = 0.74$) over a two-week interval, consistent with established benchmarks for psychological instruments (Anufriyeva et al., 2021). The WHOQOL-BREF, however, showed moderate reliability ($r = 0.66$), which may be attributed to the influence of external factors—such as socioeconomic instability or healthcare access—that fluctuate over short periods and affect quality-of-life perceptions. These findings suggest that while symptom-focused scales are relatively stable, broader life satisfaction measures may require more contextual interpretation.

Cultural and Contextual Considerations

Qualitative feedback and item-level analysis revealed that certain items—particularly those related to financial resources and health service access—were less reliable, with item-total correlations below the recommended threshold. This echoes concerns raised in previous research about the cultural relevance of imported instruments

(Weir, 2023; Flisher et al., 2005). Participants with limited literacy also reported difficulty interpreting Likert-scale options, which may have introduced response bias. These findings highlight the need for culturally adapted tools and simplified response formats to improve reliability in low-literacy populations.

Implications for Clinical Practice

The results underscore the practical utility of self-report measures in Malawian outpatient settings, especially where clinician-administered assessments may be constrained by time and staffing. However, clinicians should be cautious when interpreting results from instruments that include culturally sensitive or abstract constructs. Triangulating self-report data with clinical interviews and observational methods may enhance diagnostic accuracy and treatment planning.

CONCLUSIONS

This study affirms that self-report measures such as the PHQ-9, GAD-7, and WHOQOL-BREF can serve as reliable tools for assessing psychological symptoms and quality of life in outpatient clinical settings in Malawi. The instruments demonstrated acceptable internal consistency and test-retest reliability, suggesting they are suitable for routine use in mental health screening and monitoring.

However, the findings also underscore the importance of cultural and contextual adaptation. While symptom-focused scales performed well, broader constructs like quality of life showed more variability, likely influenced by external socioeconomic factors and cultural interpretations. These nuances highlight the need for ongoing validation and refinement of psychological tools to ensure they resonate with local populations and accurately reflect their lived experiences.

Ultimately, self-report measures offer a practical and scalable solution for mental health assessment in resource-limited settings. When used alongside clinical interviews and culturally sensitive practices, they can enhance diagnostic accuracy, support early intervention, and contribute to more holistic mental health care.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are proposed to enhance the effectiveness and cultural relevance of self-report measures in Malawian outpatient settings:

Cultural Adaptation of Instruments

- Revise and adapt existing tools to reflect local language, idioms, and cultural norms.
- Conduct cognitive interviews with patients to identify items that may be misunderstood or misinterpreted.

Simplification for Low-Literacy Populations

- Develop visual aids or simplified response formats (e.g., pictorial Likert scales) to improve accessibility.
- Provide brief orientation sessions before administering questionnaires to ensure comprehension.

Training for Clinicians and Staff

- Offer workshops on the administration and interpretation of self-report tools.
- Emphasize the importance of combining self-report data with clinical judgment and observational methods.

Integration into Broader Mental Health Strategy

- Incorporate reliable self-report measures into national mental health protocols and electronic health records.
- Use data from these tools to inform resource allocation, treatment planning, and policy development.

REFERENCE

1. Anufriyeva, V., Pavlova, M., Stepurko, T., & Groot, W. (2021). Validity and reliability of self-reported satisfaction with healthcare as a measure of quality: A systematic literature review. *International Journal for Quality in Health Care*, 33(1).
2. Dinerstein, C. (2019). Measuring the Reliability of Self-Reported Behavior. *American Council on Science and Health*.
3. Flisher, A. J., Kaaya, S. F., Butau, T., Kilonzo, G. K., Mwambo, J., Lombard, C. J., & Muller, M. (2005). Test-retest reliability of self-reported adolescent addictive and other risk behaviours in Tanzania, South Africa, and Zimbabwe. *African Index Medicus*, 4, 1–12
4. Låver, J., McAleavey, A., Valaker, I., Castonguay, L. G., & Moltu, C. (2023). Therapists' and patients' experiences of using patients' self-reported data in ongoing psychotherapy processes: A systematic review and meta-analysis of qualitative studies. *Psychotherapy Research*, 34(3), 293–310.
5. Squires, J. E., Estabrooks, C. A., O'Rourke, H. M., Gustavsson, P., Newburn-Cook, C. V., & Wallin, L. (2011). A systematic review of the psychometric properties of self-report research utilization measures used in healthcare. *Implementation Science*, 6(83).
6. Weir, C. (2023). Exploring the use of self-report behavioural science questionnaires in sub-Saharan African countries. *Doctoral Thesis, University of Stirling*
7. Zimbardo, P. (n.d.). Self-Report Measures: Psychology Definition, History & Examples.
8. Zieff, M. R., Fourie, C., Hoogenhout, M., & Donald, K. A. (2022). Psychometric properties of the ASEBA Child Behaviour Checklist and Youth Self-Report in sub-Saharan Africa: A systematic review. *Acta Neuropsychiatrica*, 34(4), 167–190