

The Imperative Of Evaluating Psychometric Properties Of Scales For Accurate Measurement: A Review

Dr Namrata Sant¹, Dr Vinay Kumar Singh², Dr Tabish Fahim³, Dr Pallavi Palaskar^{4*}

¹Ph.D Scholar, NIMS University, Jaipur Email id – namratasant@gmail.com

²Associate Professor, Department of Orthopaedics, NIMS University, Jaipur

³Professor, Department of Sports Physiotherapy, MGM School of Physiotherapy, Aurangabad.

Corresponding author ^{4*}Associate Professor, Department of Neuro-Physiotherapy, MGM School of Physiotherapy, Aurangabad
A constituent unit of MGM Institute of health sciences Navi Mumbai

Abstract

Scale development and validation represent fundamental processes in various disciplines within health, social, and behavioral sciences. However, the methodologies involved can be complex, laden with specialized terminology, unfamiliar to many, and demanding in terms of resources. Moreover, these processes are often not adequately covered in graduate education. Therefore, our aim was to succinctly outline the procedures of scale development in a clear manner, with the dual objectives of facilitating the creation of new, valid, and reliable scales, and enhancing the quality of existing ones. To achieve this, we have compiled a guide outlining best practices for developing scales to measure intricate phenomena. This guide is not a comprehensive systematic review but rather a synthesis of technical literature and insights gleaned from our extensive experience in developing or adapting various scales over several decades. We have identified three overarching phases encompassing nine specific steps. The initial phase involves generating items and assessing their content validity. Subsequently, in the scale construction phase, activities include pre-testing questions, survey administration, item reduction, and determining the scale's dimensionality. Finally, in the scale evaluation phase, dimensionality and reliability testing, as well as validity assessment, are conducted. Throughout the guide, we have incorporated illustrative examples of best practices for each step. In summary, this primer aims to empower both researchers and practitioners to comprehend the conceptual and methodological aspects of scale development and validation, thereby fostering advancements in our understanding of various health, social, and behavioral outcomes.

Keywords - Psychometric Properties, validity, reliability, factor analysis.

The Imperative of Evaluating Psychometric Properties of Scales for Accurate Measurement: A Review

The assessment tool's validity and reliability are referred to as psychometric characteristics. A scale needs to undergo thorough evaluation before it can be said to have excellent psychometric qualities, which means it is valid and reliable. Assessment of the validity and reliability of the measure is necessary in order to evaluate its accuracy and consistency. The interpretability and generalizability of the constructs being tested depend heavily on these psychometrics. Scales, clinical instruments, questionnaires, outcome measures, and specialized tests can all benefit from the application of psychometric features. The word "tool" will be used to refer to each of these categories throughout the rest of the page.

Patient-reported outcome measures (PROMs) are being utilized more and more to assess outcomes in patients with upper limb pathology as a result of the ongoing development of patient-centered healthcare models. PROMs are necessary to assess functional and quality-of-life outcomes, and they provide insight into a patient's subjective perception of their medication and health state. A PROM that is not too difficult to complete but is valid, dependable, and responsive to changes in clinical state is what makes it excellent. It is crucial to assess the psychometric qualities of PROMs before putting them to therapeutic usage in the intended population, as prior research has shown. (Ref 1,2)

The degree to which a PROM's components are pertinent to a representative of the intended construct for a specific assessment purpose is known as content validity. It addresses whether a questionnaire has enough items to cover the area of interest adequately, and whether it measures important elements of the pathology for which it is used. This is distinct from construct validity, which is established by examining the relationships between the PROM of interest and other instruments that are expected to be related, and can be measured using convergent and divergent validity approaches.

Reliability relates to the ability of scores to be repeated from one assessment to another and is commonly represented by test-retest reliability (measured by the intra-class coefficient, ICC) and internal consistency (measured by Cronbach's alpha). A PROM's responsiveness, which is a measure of its longitudinal validity, is its capacity to discern significant clinical change from measurement error. Effect size (ES) and standardized response mean (SRM) are the two metrics of responsiveness that are most frequently reported. Another crucial factor to take into account while analyzing a PROM's psychometric qualities is its floor and ceiling effects. If more than 15% of participants obtain the lowest or greatest possible score, they are deemed to be present, indicating that these patients are not likely to exhibit additional decline or

improvement. Floor and ceiling effects are harmful to a PROM's responsiveness, dependability, and content validity if they exist. PROMs are essential elements of outcome assessments in clinical practice, clinical research, and health policy (Ref 1).

The majority of PROMs evaluate symptoms, discomfort, everyday functioning, and sports engagement. (Ref 3)

Reliability and validity are important considerations while choosing measures. According to de Vet et al. (2015) The demonstration of a measure's freedom from random error is one of its psychometric pillars. Assuring the internal consistency of each item measuring the same construct in a self-report measure is one way to control for random error; another is to report test findings stability over time (test-retest reliability). (Ref 4)

An instrument's validity: it is determined by how well it measures the construct or constructs it is intended to assess. The three distinct forms of validity are as follows: (1) construct validity, which is the degree to which an instrument's scores are consistent with hypotheses, such as those regarding internal relationships, scores on other instruments, or differences between relevant groups. Also, construct validity checks that the result measure actually quantifies the construct. Convergent validity and discriminatory validity analysis are two ways to prove construct validity.

(2) content validity, which is the degree to which a measure is an adequate reflection of the construct that it intends to measure, usually determined by agreement among experts. Also, in self-report outcome measurements, face and content validity are two crucial and connected ideas. While content validity evaluates the degree to which a construct's essential ideas have been included, face validity guarantees that the construct under study is truly being observed and (3) criteria validity, or the extent to which a measure's results adequately represent the gold standard. (Ref 5,7)

Adaptability: When determining the efficacy or efficiency of a treatment, one of the most important psychometric factors is the capacity to identify both substantial and meaningful changes in the patient's state.

Time and feasibility: A well-designed outcome measure should address the total burden of completing the self-report outcome measure in addition to reliability and validity. One crucial aspect in deciding whether the outcome measure will be completed in a clinical setting—the real world—is the level of burden for both the patient and the person administering the measure. The amount of time needed to complete the questionnaire was not specifically covered in any study, although the number of items was frequently stated and might provide a physician with an approximate sense of the load. (Ref 4). Measurement properties such as internal consistency (the degree of interrelatedness among items), test-retest reliability (stability in scores over time), intra-rater reliability (relationship between scores from the same person at different times), and inter-rater reliability (relationship between scores from different people at the same time) are all included in the concept of reliability, which is the degree to which a measure is free from measurement error (de Vet et al. 2015). (Ref 6)

To ensure the validity and reliability of the measurements it is intended to measure, it is essential to evaluate the psychometric qualities of each measurement. Assurance of Dependability A scale's dependability can be established by psychometric evaluation, which makes sure the scale consistently assesses the objectives. For outcomes to be consistent across time and in various contexts, reliability is essential. Researchers must have faith that the scale yields consistent and reliable results.

1. Reliability Verification: Construct validity is one of the psychometric qualities that is evaluated to make sure the scale actually measures the idea that it is intended to measure. In order to properly draw conclusions from research findings, validity is essential. It attests to the scale's accurate representation of the desired psychological construct and nothing more.

2. Measurement Precision: Measurement precision is enhanced via psychometric assessments. In order to minimize measurement error, which has the potential to skew study results, researchers strive for high precision levels. Measuring attributes like internal consistency and test-retest dependability helps produce measurements that are more exact and accurate.

3. Generalizability of Findings: The study's findings are more broadly applicable when they possess strong psychometric qualities. Researchers can use a scale with confidence when applying it to various populations, situations, or time points if it has undergone thorough evaluation and been proven to be valid and reliable.

4. Ethical Considerations: Measurements that are inaccurate might have serious repercussions, particularly in the medical and psychological domains. It is morally required to ensure that a scale has undergone extensive psychometric evaluation in order to avoid the spread of inaccurate information or the application of actions based on unreliable data.

5. Cross-cultural applicability: When designing a scale to be used in a variety of cultural settings, psychometric assessments become essential. Cultural differences in how people react to items can affect the scale's usability and comparability across cultural groups.

6. Increased confidence in results: Rigorous assessment of psychometric properties enhances the overall trustworthiness of research findings. Researchers who have used reliable and valid measurement tools are more likely to share their findings with peers, practitioners, and others.

In conclusion, psychometric assessments are more than just a technical requirement. They are an essential part of building trust and confidence in the integrity and reliability of research. They ensure that the measurements accurately reflect the constructs being studied, providing a basis for robust scientific investigation and sound decision making. (Ref 7)

REFERENCES

1. Abbot S, Proudman S, Sim YP, Williams N. Psychometric properties of patient-reported outcomes measures used to assess upper limb pathology: a systematic review. *ANZ J Surg.* 2022 Dec;92(12):3170-3175. doi: 10.1111/ans.17973. Epub 2022 Aug 12. PMID: 35959939; PMCID: PMC10087017.
2. Palaskar, P., Attry, S., Malani, R., Gugnani, A., Kale, A., & Sant, N. (2023). Development and validation of Scale for Prediction of Ambulation in Indian Children with Cerebral Palsy- The SPAIC scale. *Medical Science*, 27(138), 1–22. <https://doi.org/10.54905/diss/v27i138/e332ms3073>
3. Gagnier JJ, Shen Y, Huang H. Psychometric properties of patient-reported outcome measures for use in patients with anterior cruciate ligament injuries: a systematic review. *JBJS reviews.* 2018 Apr 1;6(4):e5.
4. Emerson Kavchak, A. J., Mischke, J. J., Lulofs-MacPherson, K., & Vendrely, A. M. (2014). The psychometric properties of self-report outcome measures in temporomandibular dysfunction. *Physical Therapy Reviews*, 19(3), 174–185. <https://doi.org/10.1179/1743288X13Y.0000000126>
5. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights.* 2013 Aug 4;6:61-8. doi: 10.4137/HSLS11093. PMID: 25114561; PMCID: PMC4089835.
6. Du B, Lakshminarayanan M, Krishna M, Vaitheswaran S, Chandra M, Kunnukattil Sivaraman S, Goswami SP, Rangaswamy T, Spector A, Stoner CR. Psychometric properties of outcome measures in non-pharmacological interventions of persons with dementia in low-and middle-income countries: A systematic review. *Psychogeriatrics.* 2021 Mar;21(2):220-238. doi: 10.1111/psyg.12647. Epub 2020 Dec 17. PMID: 33336529; PMCID: PMC7986620.
7. Blower SL, Gridley N, Dunn A, Bywater T, Hindson Z, Bryant M. Psychometric Properties of Parent Outcome Measures Used in RCTs of Antenatal and Early Years Parent Programs: A Systematic Review. *Clin Child Fam Psychol Rev.* 2019 Sep;22(3):367-387. doi: 10.1007/s10567-019-00276-2. PMID: 30796674; PMCID: PMC6669247.