



PROFESSIONAL BOARD FOR PSYCHOLOGY

**GUIDELINES FOR THE TEACHING OF
PSYCHOMETRICS AT HIGHER INSTITUTIONS OF
LEARNING**

09/2019

INTRODUCTION

South Africa's complex and controversial political history has resulted in trends in psychometric testing and training which is imbued with discriminatory practices. In lieu of this fact, there is a need to train students on psychological tests that are applicable to the South African population. Part of the solution to this, is to revisit test selection and ways in which psychometrics are being taught at universities. This comes in light of the fact that due to the limited number of South African measures, university departments that offer training in Psychology and Psychometrics, opt to train students on international measures. Whilst these measures may have sound reliability and validity, we have to be cognisant of the cultural appropriateness and contextual realities of the individual. The teaching of psychological assessment should go beyond administering and scoring of tests, but should encompass solid psychometric and contextual knowledge in order to provide an assessment practice that is free from bias. Fostering a sense of cultural competence alongside a strong psychometric knowledge base, should be encouraged as this will enable the student to administer tests that are appropriate for a given population.

The following guidelines should assist academic departments in the training of psychometrics:

Theoretical and scientific basis

- All psychology professionals should have a sound understanding of classical psychometric theory concepts, including but not limited to:
 - Measurement scales: Nominal, Ordinal, Interval, Ratio
 - Students must be able to recognise flaws in test items such as: ambiguous items, items that may be misunderstood by other cultures, items with no correct answer, items with multiple correct answers.
 - Basic probability theory and the normal distribution
 - The importance of the mean and standard deviation
 - The meaning and properties of the different types of standard scores and how they are derived from the mean and standard deviation: Stens, stanines, T-scores, IQ-scores, Percentiles.
 - Standard error of measurement and how to implement it practically
 - Correlation – what it means and what data are required to calculate correlations
 - Significance levels of correlations vs the size of the correlation coefficient – how to interpret it
 - Reliability
 - Different types of reliability
 - How different reliability coefficients are calculated
 - The meaning of internal consistency reliability and its relationship to validity
 - The importance of variance in calculating reliability and validity
 - Understand the impact of test length on internal consistency reliability.

- Understand the impact of the number of options in a Likert scale on internal consistency reliability.
- The impact of restriction of range on calculated metric properties.
- Understand response sets and test construction techniques used to minimise them.
- Understand the difference between normative and ipsative (forced-choice) measurement and interpret test results appropriately.
- Students must be able to evaluate the quality of criterion data derived from questionnaires, performance appraisals and other performance data.
- Recognising the limitations of in-house studies (restriction of range and pre-selected samples) to determine metric properties of tests.
- Students must learn basic questionnaire construction skills.
- Students must be able to review and evaluate test manuals and documentation to determine whether a test is good enough to use.
- Students must understand the difference between bias and fairness.
- Students must understand the concept of adverse impact.
- Students must understand the concept of differential item functioning.
- Students must understand exploratory factor analysis as it relates to the validity of tests.
- Students must understand of confirmatory factor analysis as it relates to the construct equivalence of tests across cultures.
- Students must understand the factors other than the properties of the test itself that can influence the reliability and validity of measurement, and how to conduct an assessment process to optimise the probability of fair and accurate measurement.
- Students must have a basic understanding of Item Response Theory, since these tests are constructed differently and their metric properties are calculated differently.
- Students must understand that standardisation and norms are not applicable to Item Response Theory-based tests in the same way as they are applicable to tests based on classical psychometric theory.
- Students must understand when Item Response Theory-based tests are more appropriate to use than classical psychometric tests.
- While students do not necessarily need to be able to conduct psychometric research themselves, they must be able to evaluate the quality and credibility of studies in this domain.

RECOMMENDATIONS

More than one book on psychometrics should be studied, not at an introductory level. At least one book on psychometrics and assessment should be of South African origin.

Students should study the appropriate research designs and statistical analyses for cross-cultural research, especially with reference to test bias, equivalence and fairness.

Students should be sensitised to South African academic journals relevant to psychometrics, and they should at least read and discuss some articles on psychometrics published in these journals.

It is recognised that statistical software can be expensive. Universities often have academic licenses for statistical software, but once students leave the University they will probably not be able to afford the software on which they have been trained, especially if they work in a small or medium enterprise or private practice. Students should be taught how to use the basic statistical functionality in spreadsheet programmes such as Excel for simple descriptive statistics, norms calculation and correlations, and they should be made aware of free and open source statistical software appropriate to psychometrics such as, for instance, Jmetrik, PSPP and R.

PRACTICAL TRAINING

All psychology professionals should be trained in all the modes of test administration appropriate to their scope of practice – for instance: One-on-one individual assessment, group testing using pencil and paper, individual or group testing using computers with a test administrator present, remote testing using the internet with appropriate safeguards and professional practices.

Students should not be trained to rely exclusively on computerised test administration, they should be able to conduct assessments with and without supporting technology, as appropriate to the assessment being used and the respondents being assessed.

Psychology professionals who assess persons against the requirements of occupational roles, whether for vocational guidance, selection or development, should be able to analyse the requirements of the role, define the necessary competencies and relate measured dimensions to those competencies appropriately. They should be able to critically evaluate competency models and the mapping of competencies to measured dimensions to ensure that the links are at least conceptually correct and that the weighting of scales is balanced and justifiable. They should be taught to be appropriately critical of prediction models and not to simply accept them as a given.

Students should understand the importance of establishing the professional relationship before testing commences, regardless of the mode of assessment and the size of the assessment project.

Students should be able to brief respondents appropriately before assessment takes place without coaching them.

Students should be able to assess whether a person is fit to be assessed and identify persons who are not in a physical or mental condition to be fairly and accurately tested.

They should know when it is not appropriate to test a person and when alternative qualitative assessment methods should rather be used,

Students should be able to make behavioural observations during assessment where this is possible, and to make appropriate notes. They should be able to use these notes as additional qualitative information when interpreting assessments.

Students should be able to obtain informed consent, considering the age and background of the respondent, their capacity to consent, ensuring that consent is fully informed, obtained in an appropriate manner and that a record of the consent is obtained. Students should be aware of the need to obtain consent that allows the results to be used for validation research. Students should be able to design and/or modify consent forms appropriately if existing forms do not meet the requirements.

Students should obtain practical experience in running assessment projects of various sizes. They should be able to select the test battery appropriate to the respondents and the purpose of the assessment, to budget for the project, coordinate the logistics, delegate appropriately and deliver the finished reports on time and up to standard.

Students must be taught appropriate interviewing skills for their scope of practice and the respondents they need to assess. They should be able to do an intake interview, an interview to verify the accuracy of information obtained from psychometric tests, and a final feedback to the respondent, and a feedback interview to a stakeholder (Who could be a teacher, parent, employer or other professional). The tone and level of the feedback must be appropriate to the intended recipient and appropriate interpretive aids must be used, without distorting the results.

All psychology professionals should be able to produce an integrated report based on several different assessments, also integrating qualitative information obtained from interviews and other sources. The interviews must be conducted in an appropriate style for the recipient of the feedback, and must be clear and understandable to the intended recipient. Students must be taught not to rely exclusively on computer-generated reports, although these may be utilized in the production of an integrated report.

Students should be taught how to maintain appropriate security over personal information, assessment results and assessment materials and equipment. They should be able to do this for electronic assessments, pencil and paper test materials and apparatus tests where appropriate.

TRAINING ON SPECIFIC TESTS

It is recognised that accreditation training for some tests can be very expensive. However, students should understand the need for accreditation training when test distributors require this.

Students should be exposed to a range of measures and should be taught to select tests appropriately, considering several alternatives and choosing the most appropriate one for the purpose and the respondent group.

Students should be exposed to tests that measure all the important constructs relevant to their scope of practice, preferably more than one test per type of construct (For instance, more than one ability battery, more than one personality test and more than one interest questionnaire, etc.).

RECOMMENDED READING LIST:

General Psychometrics

Allen, M. J., & Yen, W. M. (1979/2002). *Introduction to measurement theory*. Prospect Heights, IL: Waveland Press.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: Holt, Rinehart, & Winston.

Foxcroft, C., & Roodt, G. (Eds.). (2005). *An introduction to psychological assessment in the South African context* (2nd ed.). Goodwood, Cape Town: Oxford University Press.

Nunnally, J.C. & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw Hill

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Routledge.

Rust, J., & Golombok, S. (2009). *Modern psychometrics, third edition: The science of psychological assessment* (3rd ed.). New York, NY: Routledge.

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons.

Factor Analysis

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist*, 34, 684-718.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modelling* (3rd ed.). New York, NY: Routledge.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Modern Measurement Theory

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.

Hambleton, R. K., & Swaminathan, H. (1984). *Item response theory*. Boston, MA: Kluwer-Nijhoff.

Hambleton, R., K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Reise, S. P. (2009). The emergence of item response theory models and the patient reported outcomes measurement information systems. *Austrian Journal of Statistics*, 38(4), 211-220.

Rupp, A. A., & Zumbo, B. D. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and Psychological Measurement*, 66(1), 63-84.

Reliability

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34, 177-189.

Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage.

Streiner, D. L. (2003a). Starting at the beginning: An introduction to coefficient alpha internal consistency. *Journal of Personality Assessment*, 80(1), 99-103.

Streiner, D. L. (2003b). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217-222.

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53(1), 33-49.

Test Construction

McGrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment*, 85(2), 112-124.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.

General Statistics

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th ed). New York: Holt, Rinehart & Winston.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Cross-Cultural Assessment and Equivalence Testing

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European psychologist*, 1(2), 89.

Van, de Vijver, & F., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. *Cross-cultural research methods in psychology*.

Van de Vijver, F., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13(1), 29.

Van de Vijver, F., & Tanzer, N. K. (1998). Bias and equivalence in cross-cultural assessment. *European review of applied psychology*, 47(4), 263-279.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, 12(3), 1-26.

Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language testing*, 20(2), 136-147.

Best Practice Guidelines

Muñiz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12(3), 206