

Evidence-based Psychological Assessment

Dragos Iliescu¹

¹*Department of Psychology, University of Bucharest, Romania*

Keyword: Psychological assessment, evidence-based management

Abstract: Testing and assessment is possibly the single most influential contribution of psychology to the repertoire of scientific methods. But testing and assessment has evolved in meandering and not always straightforward ways during the past 100 years, combining today a number of extremely modern approaches, but being also tributary to historical artefacts that cannot be considered up to par with scientific requirements. In the same way in which we promote today an evidential basis in those areas of psychology that are dedicated to interventions (e.g., evidence-based psychotherapeutical interventions, evidence-based management), we should promote an evidential basis in psychological testing and assessment. This lecture will concentrate on the definition of evidence-based assessment, will discuss the different ways in which evidence-based assessment may be approached in clinical, educational and work psychology, and will finally focus on a number of utility analyses related to evidence-based psychological assessment, as opposed to more traditional methods of assessment that do not always have an empirical basis. This article was presented as Keynote Lecture for the International Conference of Psychotechnology (ICOP) being held in on 5-6 September 2018 at Bina Nusantara University (BINUS) - Alam Sutra Campus, South Tangerang, Indonesia.

1 INTRODUCTION

Evidence-based psychological assessment (EBPA) is psychological assessment guided by the use of both current theory and current scientific research. Theory and research are used in order to ground decisions on the whole assessment process, from the selection of the constructs that are used, the selection of measures or mix of measures for those constructs, the administration, scoring and interpretation of those measures, to defensible decisions made based on the gathered evidence.

EBPA was adopted initially by medicine and has migrated towards psychology on this route, through clinical psychology; it is however used today in all fields of psychology. It is less arbitrary and subjective than the traditional approach to assessment, and has been reported to be highly efficient by both empirical studies and systematic reviews (Dawn, Légaré, Lewis, Barry, Bennett, Eden, Holmes-Rovner, Llewellyn-Thomas and Lyddiatt, 2017). EBPA is now a mature scientific stream, that has generated impactful practices and an important body of literature (Antony and Barlow, 2010; Antony, Orsillo and Roemer, 2001; Nezu,

Ronan, Meadows and McClure, 2000), that guides psychological assessment in a number of fields.

EBPA is strongly empirical, i.e., it assumes that the lack of empirical evidence for validity in a specific context or for a specific use degrades a method to the point that its usage is questionable. However, EBPA does not exclusively rely on empirical evidence, but acknowledges the fact that the assessment process is a decision-making task where professional judgment prevails. As an effect, it combines the best available evidence with the preferences of the client, and with professional expertise (Bornstein, 2017).

As we see, strong evidence for validity is an important requirement. This relates to the psychometric characteristics of the measure or measures used. It is improper to use in an EBPA approach measures with a known weak reliability and validity evidence. Similarly, while absence of evidence is not evidence of absence, lack of strong evidence of validity degrades the legitimacy of a measure to the point where it should be avoided.

Equally important, evidence should be strong and should be contextual to the usage it is given (Hunsley and Mash, 2007). The criteria for strength of evidence in EBPA are similar with those for

clinical decision-making, with systematic reviews and meta-analyses providing the strongest evidence, randomized control trials and experimental studies providing strong evidence, and with cohort studies, single case designs, case studies and expert opinions providing consecutively weaker evidence. Many evidences for the qualities of a specific measure are generalizable, but it is important to have evidence specific to the context the measure is applied in (i.e., the characteristics of the administration context and decision context, the specific population, language etc., it is used), or a strong and legitimate argument for the fact that, even lacking such direct evidence, the general evidence should apply.

Evidence is not limited to the psychometric characteristics of a measure. Oftentimes assessment is conducted with several measures, either in a multiple-hurdle process, or by combining measures based on statistical or judgmental decisions. Evidence should exist about the interaction of these measures, the incremental validity of each, the manner in which they could be combined in decision-making, and evidence should be specific to the exact manner (statistical or judgmental) in which the results of the several used measures are actually combined. We feel compelled to note here the wide preference of psychologists in general, and clinicians especially, for judgmental decisions (i.e., decisions based on clinical judgment, personal experience and ‘flair’) (Garb, 2005). The evidence points very much against such decisions (Ægisdóttir, White, Spengler, Maugherman, Cook, Nichols, Lampropoulos, Walker, Cohen and Rush, 2006; Davis, Mazmanian, Fordis, Van Harrison, Thorpe and Perrier, 2006) and research has shown rather clearly that psychologists, like humans in general, are prone to bias have limited awareness for this fact and are because of this influenced in their decisions by personal beliefs, heuristics and sometimes outright biases.

Also, evidence should be provided on the recommended cut-off scores and the likely margin of error in scores in general and in cut-off scores especially. Such evidence could be related to error brackets resulted from the reliability of the test or assessment system (e.g., standard error of measurement or standard error of prediction), but could also be based on sensitivity-specificity analyses outlining the false positives and false negatives likely to appear in a decision based on that test in that context.

However, in EBPA strong evidence for validity on behalf of a specific measure, showing good psychometric qualities and a strong relationship to a target criterion, are not enough. The process

combines this “best available evidence” with client preferences (Hunsley and Mash, 2007). This means that the characteristics of the context in which the assessment is conducted, such as client characteristics, situational characteristics etc., are equally important. This acknowledges the fact that there is no such thing as “the best measure”, but that the adequacy of a measure is an interplay between psychometric characteristics and evidence and the constraints of the situation. There are many reasons why a measure is not applicable to a specific client: language barriers, physical barriers, outright rejection, lack of face validity. Quite aside from reasons for applicability, the clients also have preferences, and these should be taken into account. As a result, EBPA does not necessarily encourage usage of the psychometrically strongest measure, but rather the usage of the most adequate measures.

Finally, professional judgment, or ‘clinical expertise’ also play an important role in EBPA. The EBPA process is ultimately a problem-solving decision-making task, and like any problem-solving and decision-making task it is profoundly iterative in nature. The psychologist will formulate hypotheses, and test these hypotheses by generating data, interpreting data and integrating data from multiple sources, oftentimes encountering data that are incomplete, inconsistent or both. In such cases, the capacity of psychologists to bring their experience to bear, through professional reasoning, is critical for a useful decision.

2 FUTURE DIRECTION

We urge psychologists to acknowledge that it is impossible to have useful professional reasoning, and as a result it is impossible to encounter EBPA without strong professional expertise. Such expertise is certainly related both to the subject matter on which assessment decisions need to be made (i.e., autism spectrum disorders, learning disabilities, or job recruitment and selection), and to testing expertise.

Testing literacy is often low in test users, and this impacts the quality of their decisions considerably, even if they are experts in their field. A higher testing literacy leads to more competent test users and competent test users conduct EBPA. It is rather difficult to describe a “competent test user” – arguably this is impossible in absence of a context in which test usage would be applied. A test user may be very competent in the context of job selection assessment, and less competent in educational and

clinical assessment. Competence certainly *is* contextual, it requires a substantive knowledge of the field and hands-on experience with assessment in this field. At the same time, long-lasting hands-on experience may be tainted by older practices that were not informed by science, or that are not informed by the latest state of science. In many countries around the world, the community standard for clinical assessment, even for high-stake decisions, is based on projective techniques. The psychometric and validity evidence for projective techniques is such that they should be generally avoided, or at the very least shunned for high-stake assessments. Still, long practice and generations of psychologists learning in mentorship relations about assessment, have ingrained such practices to the point that they are accepted without critical thinking. As a result, long-lasting hands-on experience is not necessarily an indicator of competence.

The International Test Commission (ITC) has developed an important document entitled “*The ITC Guidelines on Test Use*” (International Test Commission, 2001). This document outlines a number of general characteristics (i.e., characteristics that are not context-related) of a competent test user. We advance that EBPA is impossible in the absence of these test user characteristics.

The document states that “a competent test user will use tests appropriately, professionally, and in an ethical manner, paying due regard to the needs and rights of those involved in the testing process, the reasons for testing, and the broader context in which the testing takes place” (p. 6). These Guidelines have been developed in such a way as to be applicable internationally. They explicitly recognize that many contextual factors may affect how these precepts are applied in practice in a specific culture and/or country, and urgent test users and regulators to consider social, political and historical differences, as well as specific laws and regulations when applying these guidelines.

The ITC Guidelines have 2 parts. The first part describes competent test users as those users who take responsibility for ethical test use by acting in a professional and ethical manner, ensuring they have the competence to use tests, taking responsibility for their use of tests, ensuring that test materials are kept securely, and ensuring that test results are treated confidentially. The second part describes competent test users as those users who follow good practice in the use of tests, by evaluating the potential utility of testing in an assessment situation, choosing technically sound tests appropriate for the situation,

giving due consideration to issues of fairness in testing, making necessary preparations for the testing session, administering the tests properly, scoring and analyzing test results accurately, interpreting results appropriately, communicating the results clearly and accurately to relevant others, and reviewing the appropriateness of the test and its use.

REFERENCES

- Ægisdóttir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Cook, R.S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G. and Rush, J.D. (2006) ‘The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction’, *The Counseling Psychologist*, vol.34, pp. 341-382.
- Antony and Barlow (2010) *Handbook of Assessment and Treatment Planning for Psychological Disorders*, 2nd Edition. Guilford Press.
- Antony, M.M., Orsillo, S.M. and Roemer, L. (2001) *Practitioner's guide to empirically based measures of anxiety*. New York: Plenum.
- Bornstein, R.F. (2017) ‘Evidence-Based Psychological Assessment’, *Journal of Personality Assessment*, 99(4), pp. 435-445. doi: 10.1080/00223891.2016.1236343
- Davis, D.A, Mazmanian, P.E, Fordis, M., Van Harrison, R., Thorpe, K.E. and Perrier, L. (2006) ‘Accuracy of physician self-assessment compared with observed measures of competence: A systematic review’, *Journal of the American Medical Association*, vol. 296, pp. 1094-1102.
- Dawn, S., Légaré, F., Lewis, K., Barry, M.J., Bennett, C.L., Eden, K.B., Holmes-Rovner, M., Llewellyn-Thomas, H. and Lyddiatt, A. (2017) ‘Decision aids for people facing health treatment or screening decisions’, *The Cochrane Database of Systematic Reviews*, vol.4. doi: 10.1002/14651858.CD001431.pub5
- Garb, H. (2005) ‘Clinical judgment and decision making’, *Annual Review of Clinical Psychology*, vol. 1, pp. 67-89. Nezu, A.M., Ronan, G.F., Meadows, E.A. and McClure, K.S. (2000) *Practitioner's guide to empirically based measures of depression*. New York: Kluwer Academic.
- Hunsley, J. and Mash, E.J. (2007) ‘Evidence-based assessment’, *Annual Review of Clinical Psychology*, vol.3, pp. 29-51. doi: 10.1146/annurev.clinpsy.3.022806.091419
- International Test Commission (2001) ‘International Guidelines for Test Use’, *International Journal of Testing*, 1(2), pp. 93-114.