AN ETPG WHITE PAPER CHOOSING PSYCHOLOGICAL TESTS FOR USE IN HEALTH SETTINGS: QUESTIONS TO ASK

Key Authors: Natasa Ko, Camelia Ionescu

Authors, publishers, and assessment users create a partnership to ensure responsible test use. The European Test Publishers Group (ETPG) works with thousands of test users internationally and these are the most common questions asked by users about tests in healthcare settings. ETPG seeks to provide answers and suggest how to identify the best test for your purpose. Different countries run different professional, legal and scientific guidelines, rules and policies for test use and supply. This paper attempts to describe general navigation principles but these should be adapted to meet the requirements of such local guidelines and legislation.

Underlying guiding principles for the use of any test are:

- *Responsible test use is a partnership between publishers/distributors and users.* ***Publishers/distributors*** *should provide enough accurate information for test users to make informed choices about which tests they need and how to use them.* ***Users*** *should use them professionally, taking time to learn how to use each test they find appropriate for their work.*
- *You should know why you are using a test in a particular circumstance with a particular client and/or patient. Different tests are better or worse for different purposes.*
- *The client or patient's needs come first. To use a test without appropriate knowledge affects test takers lives adversely.*

GENERAL QUESTIONS ON CHOOSING TESTS

I do not have time to follow a systematic approach when choosing tests. What could I do to ensure I still stick to best practice?

Sometimes, choosing a test can be time-consuming. To learn important information in a time-effective way, look at the publishers'/distributors'/authors' website description of the purpose and scope of the test. Refer to the test manual or search scientific databases for relevant publications.

A suggested minimum approach is to check for transparent information about:

- The theoretical framework/model behind the test,
- Psychometric qualities of the test
- Relevance, age, and robustness of test norms,
- Whether a test meets the requirements set by local professional or scientific organisations (for instance the criteria of the EFPA test review system).

This is the overarching process that will answer many of the questions in this white paper. We will look at these in more detail.

What should test choice be based on: technical quality or popularity?

Popularity is not necessarily a good guide. What matters – among other issues – are the theoretical background, the reference group, and validity and reliability information, etc. You can look these up in the test manual or in scientific publications. You may have test preferences but some very popular tests are designed only for certain purposes or may not have kept up with latest research findings and thinking.

New tests may not be cited enough in academic studies– should I still include them in my practice?

Research data supporting test validity and reliability in the manual of a well-developed test should be enough to evaluate a test. If external research data is available, the quality and relevance of the publication in which they appear is important. Then check periodically for new associated research, which is also advised for renowned tools. Psychology and measurement studies are changing so it's important to keep up with best practice.

Some tests are very popular despite there being no solid available proof evidence that they work (psychometric, sampling, norming data) beyond personal experience. Some are free and it is claimed they add some value/output since they are very time-efficient. Should I use some of these?

Subjective experience ('I've used it and it works') can be very misleading. If there is no objective proof that something works, you cannot tell what an assessment is doing. Good psychological tests conform to many scientific and psychometric criteria. Anything that fails to comply with these may be *a kind of measurement tool*, but the validity of the results can not be professionally judged. Using tools which meet scientific criteria protects you from unwanted consequences with ethical or legal impacts.

On the additional point, it is not easy to find valid and reliable tests that are freely available. Many are loss-leaders for commercial firms offering other services and may be technically weak. However, you may find a relevant, strong test published in a scientific journal that was made available by the author (mainly for research purposes). If your purpose differs from a research one, then you might find that adding norms to your test and ensuring that it meets local standards becomes expensive rather than free which makes the test no longer free.

Some of the tests I would like to use have not been assessed (therefore not approved) by the local psychological authority – can I still use them?

As noted at the beginning of this paper, local arrangements for test supply vary from country to country. Some countries do not have a test review system; some do not require a test to be approved by a professional organisation. Some, on the other hand do. Even if there is a formal arrangement for test approval, not all tests will have gone through the process by the time a user would like to use it. If there is no such system, look for and evaluate the information we suggested at the beginning of this white paper; also evaluate the sorts of technical information the publisher makes available.

If there is a formal review system, ask the publisher if and when they plan to submit the test for local approval.

Why are tests so expensive?

We believe good, scientific tests aren't expensive, given their status as scientific instruments and the work and finances that go into ensuring their accuracy, and the life-changing decisions their outputs

underpin. The cost of bad decisions partly based on evidence from technically weak tests or no assessment data at all can be very high.

Developing a good test is a costly and time-consuming process requiring the contribution of many professionals (academics, practicing psychologists, data collectors administering the test to hundreds or thousands of individuals as part of the norming process, data scientists, programmers, writers and graphic designers). Test versions are sometimes being developed and tried out in recurring cycles for years in order to ensure scientific proof that they work (are reliable and valid).

Investment in new tests is critical and the vast majority of test publishing houses do not receive public financial help.

The vast majority of free tests do not reach local quality standards and may cause significant problems for you and your client.

I sometimes make copies (physical or digital) of tests, what are the risks?

Do not do this unless you have contacted and had written approval from the copyright owner. Tests are protected by copyright; users therefore are required to purchase original test material. Copyright is a property right and copying without authorisation is theft. The worst-case scenario is that you may be sued if you copy without permission.

Copying reduces the amount of funds available for test development and research and supports Copying can affect the integrity of the test material, especially if the graphic details (colours, shapes, and lines) become altered by the copying. This may result in wrong diagnosis and decisions about your client.

How many tests should I have available and be able to use?

This depends on the mix of your workload and the sorts of clients you work with. The more divergent these are, the more likely that you will need more and more specific tests. On the other hand, relying solely on the results of one test is not recommended. In the clinical arena, in addition to the large multi-psychopathology tests, there exist instruments for personality, distinct disorders (depression, anxiety, alcoholism), risk assessment (suicide), general functioning, adaptive behaviour, and there also exist diagnostic interview schedules.

Why am I not allowed to use some tests?

As must be clear, test user knowledge and skill are central to correct use of tests. For this reason, supply of some tests is limited to those who have certain initial, professional, and in-service training, to protect users from misuse and wrong decisions.

ARE SPECIFIC TESTS SUITABLE FOR MY CLIENTS/PATIENTS?

Some tests do not yet have a local standardised version in my country, and some have some unofficial/informal version, often without local norms – can I use them?

It depends on the test whether it requires adaptation to local norms or doesn't: the majority of them do. The translation and adaptation process of any unofficial/informal versions is often unsupervised and any collected norms may not have been checked by the copyright owner, so they may not have adequate validity and reliability. The copyright owner usually gives an adaptation and distribution licence to a reliable and authorised local test distributor trusted to have the necessary expertise to carry out the process.

When using tests in my practice I frequently think that some items are not appropriately adapted to my culture or language – can I suggest different items or explain the items to my clients?

You might have spotted a problem; that depends on the origin of the version you have. If it is a licensed version, you can have more trust in the translation or the item content as the adaptation was done by the test distributor. Yet, even in this case, it might seem that certain items look awkward or inappropriate. This may depend on subjective factors, or it can be that you spotted an item that is still the best approximation of the original item (so it is a compromise made by the adaptor/translator) or best reflects the range of the latent trait. However, If the test is unlicensed, items are not necessarily translated or adapted well. In the majority of psychometric tools, data sampling is representative and takes into account specific populations. If you have concerns about this, you should check the norms in the test manual.

In either case, you should not suggest a new item ad hoc. To the extent the manual allows it, you can explain the item to the subject. If many of the users think some items could be replaced and then they do so, this would eventually invalidate the standard administration, scoring and interpretation and may risk violating copyrights of authors and/or translators.

If you perceive that a minority subject finds some items confusing, give as much explanation as the test manual allows, or you may attempt the assessment with a *culture-fair test*.

DETAILED TECHNICAL ISSUES

Can I trust tests that have no validity scales? How can I prevent distortion in clinical context?

These scales are important to determine the validity of the subject's responses (to determine biased or random responding). It can help if you emphasise to the test-taker the need to pay attention and to give objective, candid responses but this does not necessarily increase validity of the responses. If a test does not include such scales, you need to rely on other clinical information to obtain supporting evidence for a client's results.

How do tests account for any limitations, caveats, or uncertainties in the assessment or evaluation process?

This is a multifactorial issue. Test manuals usually prescribe the limitations or caveats of the use of the test results. In addition to considering those described in the test manual and the well-known basic psychometric principles, clinicians must pay special attention to the sources of measurement error, score confidence intervals, meaning of T-score ranges (standard deviations), interpretation of the various T-score ranges, correspondence among the various standard scores, interpretations and inferences allowed for certain scores, limitation in predictions, and concordance with constructs of other tests, etc.

How can I check that a test aligns with current research and best practices in psychological testing and assessment?

If your test version is the latest one published (by the copyright owner), then the model behind it, and its scale structure should be considered valid. If you have an earlier version, its validity depends on how much radical change (scientific models, construct, scale structure, item content, etc.) took place between the two versions. You can also overview the current trends in test use or the validity of the version you have in the literature.

ISSUES IN USING TESTS WITH A CLIENT

Clinical tests take too long to be administered – what can be done?

Most tests aim to gather the most valid assessment data within the shortest time and length possible. If there is no shorter administration option offered in the test manual, cutting the length of the test or the administration time on an arbitrary basis is not possible without compromising validity.

How can I monitor treatment/progress in clinical context using tools?

Pre- and post-assessment with the same tool can be done. You will need to pay attention to the proposed time interval between testing sessions. Some tests are inappropriate for use in this way and any exposure to test items will tend to consolidate them in a client's memory, affecting their responses. The test manual should provide advice on these issues.

What dimensions of personality as measured by tests impede the therapeutic process and should be assessed beforehand?

Although many aspects of personality can affect therapeutic progress or personal growth, we can place more weight on the *motivation* to change, *optimism*, or, for instance, *self-efficacy* (the belief that joint therapeutic work will be successful), among others.

What makes a test suitable for case formulation of different therapy approaches?

Case formulation involves summarising client information, referral questions, important life events and identifying core strengths but also limitations, in order to inform recommendations, diagnosis and prognosis. Some therapy approaches exclude test usage, but good tests can deepen your professional understanding of a case. The majority of tests available for clinical settings can successfully be used in case formulation: any therapy approach can benefit from an objective entry point; any client can start the self-development process by understanding how his/her/their inner world is currently organised.

How can I match best the (different) clinical descriptions used by various tests?

This is the responsibility of test users who need to build on professional knowledge by using different sources of information, integrating various test results, comparing information from different sources, and coming to reasonable conclusions.

The increased use of computer-generated interpretive reports highlights this point. Users must be aware that they must not solely rely on such a report alone, and that it is just initial raw material of for the interpretative process. This does not imply, however, that you need to possess all knowledge: you can look up and rely on the research and the body of scientific knowledge available about the results, and the test and other interpretive manuals. The process requires clinician objectivity, common sense and involves a high level of responsibility.

What part of the results can I share with the client? What level of debriefing is allowed?

Test takers have the right to receive information about the purpose of the testing, the intended use of test scores, and about the test results. The scores and interpretations should be expressed in terms that can be understood by the client; in some countries users are trained NOT to share numerical information as this has proved open to misinterpretation. Clinical tests are a specific issue – as opposed to, say, tests used in businesses for counselling and development - as they are typically intended to be used by health professionals. Clinical reports are usually not handed over to the clients, as reading their content may cause harm to them. Feedback can be given verbally to the

client in a careful, positive, and supportive way. You should check standard practice in your country, profession, and institution.

Tests have been administered and these are the results. What steps should I follow? How do I integrate them all?

There are several factors that should be taken into consideration before proceeding with the interpretation: not least the sorts of test technical data we have highlighted throughout this paper. Other important information includes signs of interest and motivation in the examinee, the effects of individual characteristics (e.g. demographic variables, employment status, linguistic preference) on test results. It is crucial to factor in test taker interest, motivation, and effort.

This partially highlights the importance of establishing and maintaining *rapport* with the test taker. If, despite your efforts, and due to the examinee's state, you perceive that he/she is not in his/her best *mental, emotional,* and *health condition*, the evaluation should be postponed.

Most test manuals contain information about how to proceed with the results. However, they do not always contain every single step and detail, as the professional interpretation may go beyond the scope of a test manual. Integration of the various test results is mainly the competency of the psychologist (see earlier in this paper).

How do I deal with positive replies to critical items?

Always check for the bigger picture. Critical items are *mainly informative of the topics the subject chose and do not necessarily inform about the severity of the indicated problem area*. They can not be validly interpreted if their corresponding parent scales are not elevated. E.g. If these are the Critical items of the most recent MMPI versions, please check a renowned interpretive manual for how these should be interpreted.

How can I know that the tests and measures administered, scored, and interpreted by another practitioner are safe to use to continue the treatment plan?

Objective tests are the ones that make standard administration, scoring and interpretation possible. Therefore, it can be expected that if this process is done by two highly qualified professionals in the standard way, they should arrive at comparable score interpretations. Some tests (*Projective tests* for instance) on the other hand, are generally less standard in administration, scoring and interpretation, therefore the outcomes and conclusions by the two professionals may be somewhat/largely discrepant.

How can a test's results impact the referral question or hypothesis?

If the results come from objective tests and the administration-scoring-interpretation was standard and completed by a highly qualified and competent professional, they can be considered valid and reliable, therefore can be trusted. So, they may influence or modify the referral question or the hypotheses.

How much can a test's results relate to the client's presenting problem, diagnosis, treatment plan, and prognosis?

This depends on the *purpose of the test* which is specifically detailed in the test manual (if it provides diagnosis, helps in treatment planning, and informs about the prognosis). The scope of the test is crucial, for instance, whether *it provides diagnosis* or the results *only contribute to* the formulation of a diagnosis. Tests have a "utility" aspect in that they are intended to help the professional to understand and treat the client, to improve his/her overall condition.

What should I do when test results contrast with previous or concurrent assessments of the client?

Non-concordant results may be due to many factors. How much time passed between the two test-sessions? Was the administration-scoring-interpretation standard with all tests? Was the client in similar condition, was he/she similarly motivated at the testing occasions? Are the tests valid, legally licenced, well translated or adapted? Do they measure similar or entirely different constructs? Is there enough evidence for valid score interpretations with the chosen instrument or chosen use, etc.? Non concordant results can also be due to the fact that two tests or scales actually measure two or more different constructs, yet the professional believes that these constructs are overlapping or are the same. For instance, when measured with precisely formulated pools of items, it can be demonstrated that anxiety and fear are two different constructs, yet the clinician believes that they are identical therefore should score similarly. An even more subtle confusion can be seen when it comes to *empirical correlates* of some scales. For example, correlate *subjective depression* is different from correlates that refer to *others perceiving the subject as depressed* (the self and the others' view difference).

What are the ethical recommendations when it comes to emergency or critical situations (e.g. a potential suicide attempt) when the clinician has to make test results available to a third party to prevent harm-causing actions, thus breaching privacy & confidentiality with the client?

Protection of subject privacy in psychological measurement is a basic principle. The subject needs to give agreement prior to testing to enable the clinician to disclose (specified) test information to other professionals. Privacy can be breached (in emergency cases) only after a proper risk assessment, to prevent harm and ensure protective measures, limiting disclosure, as much as possible, to essential information, as defined by national laws. Legal aspects/local laws must be taken into account when faced with a decision prioritising safety over autonomy and should be preferably included in the (recommended) prior test agreement, above mentioned.

Can the results of a test reflect a client's strengths, weaknesses, preferences, personality, coping style, and cultural background?

Tests provide different scale results *the content of which* may or may not cover the topics above. If there are no relevant, straightforward scales, pl. consult the test manual or interpretive manuals on *how to infer* strengths and weaknesses, coping, or personality etc. from the results. If such guides are not offered, the test information may not be relevant in these respects, and assumptions about these can not be made.

Can a test inform a client's goals, needs, preferences, and expectations for therapy or intervention?

The client's intervention-related or therapeutic expectations may be of any kind and can likely be observed in his/her treatment "attitude," the level of cooperation shown during the therapeutic progress. This is influenced by many factors: treatment-seeking motive, trust in the therapist, compliance, and the readiness to disclose personal information, optimism, self-confidence, and self-efficacy, etc.

Can test results be communicated more widely than to the test user?

Test users must develop some skills in communicating the implications of test results to those who have the right to receive them. The feedback should be clear, *constructive*, considerate, tactful, meaningful, and honest, in a language understandable by the examinee. Some computer-administered tests provide different reports for different audiences =using different, appropriate language and formats for each.

Can a test provide specific, measurable, achievable, relevant, and time-bound recommendations for action or intervention?

Tests differ in what they offer for treatment and intervention. Particular approaches are described in test manuals, sometimes in detail, some less so. Some tests analyse their results and suggest courses of action though these suggestions need interpretation by a qualified psychologist. measurement.

How can test results be used while protecting the confidentiality and privacy of the client and their data?

A test user is responsible to keep all sensitive data: personal information, test related personal data, the response forms and test report, in a secure place or in a secured way. Personal test data, and test results can not be disclosed unless an authorised professional or authority is entitled to see and evaluate them. Even if this is the case, disclosure of such information can take place in accordance with any applicable privacy laws.

Is it possible to justify any deviations from standard procedures or protocols in psychological testing and assessment (e.g., use of non-standardized or adapted instruments, administration by telehealth)?

Some tests allow minor accommodations during the *administration of the test.* These options are detailed in its manual (different modes of presenting the test, or its instructions, the order of the subtests, or of responding). The accommodations should be noted on the record form and taken into consideration when the results are interpreted. Any different or greater deviation than these may not be allowed by the manual. This involves using a test that has not been adapted (yet it should have been) or having an interpreter *translating the verbal items* to the examinee speaking a different language from the language of the assessment. Telehealth administration may be permitted by the manual or described as an additional mode of administration on the website of the test distributor.

Can the results of tests and measures contribute to ongoing evaluation and quality improvement of psychological testing and assessment services?

Tests are continuously being developed, not only scientifically, but also by making administration, scoring and interpretation more manageable while maintaining the focus on the complexity of examination and on high validity. Before designing a more advanced version of a test, developers gather users' field experiences and review scientific data from publications of the current version. This helps them to improve tests not only theoretically (based on the e.g. latest cognitive models or diagnostic categories, etc.) but also building on the accumulated test results and test users' practical experiences. See also points 22 and 28.

Is it ok to use AI to integrate my test results?

This is an evolving area so there is as yet no clear answer. A good starting point to think about it is that when professional judgements, reports and opinions are expressed in a verbal or written format, the given professional assumes responsibility for his/her statement. If AI generated opinions, statements are included, who is supposed to assume responsibility for these? This seems to be a psychological ethical and a legal issue. Even if the clinician only builds on and then alters the AI generated report, he/she can hold the belief that part of the responsibility does not lie with him/her.

My client was assessed by an illegal (foreign) test as part of an official examination for evaluating parental fitness (adoption case). I am aware that the test results are invalid and my client's request for adoption was declined based on the results. How can I help him/her?

If you want to help, first you will need to gather evidence that the version used does not have an official adaptation and distribution licence. A written proof can be obtained from the foreign licensor that can be submitted to the agency responsible for the assessment and the decision. It can be expected that this might overturn their decision (even if the results only played a smaller part in the decision). For this process, involvement of a legal representative may be important.

ABOUT THE ETPG

The European Test Publishers Group (ETPG) is a group of psychological test publishers, all based in Europe. It was formed in 1991 to: *'help create an intellectual and commercial environment which values scientific measurement in psychological testing. It supports its members and works with test customers, users, developers, researchers, academics, and policy makers to achieve this.'*

ETPG's annual conference focuses on new developments in psychological testing, looking at ways our industry can meet changing needs. Our white papers on key testing topics contribute to scientific and professional conferences as well as relevant policy proposals. We welcome discussions on ideas will improve testing's impact on European society.

ETPG works with other associations (the European Federation of Psychological Associations [EFPA], the International Test Commission [ITC], the Association of Test Publishers [ATP] among others) to continually improve testing practices. Its annual conference is a major forum to discuss new test developments and the needs of European test users. ETPG gathers together individuals and companies with huge experience and knowledge in all aspects of psychological testing.

The ETPG's mission is to improve European Testing by:

- raising standards in test design and use;
- progressing creative developments in testing;
- promoting professional use of tests internationally;
- developing fair systems of copyright protection to the benefit of all parties;
- increasing and promoting the benefits of applied psychology to European citizens;
- developing links with test industries outside Europe to raise overall standards.

The group's values focus on:

- Innovating
- Social Engagement
- Being evidence-based
- Using scientific methods to develop products and services.
- Being open and transparent
- Adhering to professional standards