

## **ATP SA RESPONSE:**

### **Re-classification of Psychological tests on the current list of classified tests (document for discussion and consultation)**

Herewith the response of the Association of Test Publishers of South Africa (ATP SA) in reference to the Professional Board for Psychology's request for feedback on the re-evaluation / re-classification of tests (dated 10 April 2006, Ref: 18/11/B). ATP SA appreciates the opportunity provided to give input on the Board's proposal for the re-evaluation / re-classification of tests. We would like to emphasise that we, as test publishers, undertake to act professionally, ethically and in a responsible manner.

ATP SA believes that, in working together with the Board and Psychometrics Committee, we can make the process of classification of psychological tests easier.

## Table of Contents

Background and History of the Association of Test Publishers .....	1
General Standpoint of ATP SA Concerning Test Classification in South Africa.....	4
ATP SA's Commentary on the Board's Reclassification Process Proposal .....	6
Reasons for the need to re-classify psychological tests .....	6
Historically classified tests that have never been reviewed and evaluated ....	6
The need to accommodate newly created registration categories .....	8
Ensuring consistency, clarity and legal compliance .....	9
Ensuring that tests are adequately supported and updated .....	10
Proposed process for the re-classification of tests .....	10
Comment phase .....	10
Review and classification .....	10
Order of priority .....	11
Payment .....	12
Appendix A: Global Trends and the Regulation of Psychological Tests .....	13
Appendix B: Grading of Psychometric Tests and Qualification Levels of Test Users .....	14
Appendix C: EFPA Review Model for the Description and Evaluation of Psychological Tests.....	15
Appendix D: Competencies of Test Administrators .....	16
Appendix E: BPS Test Review Process .....	17

## Background and History of the Association of Test Publishers

The Psychometrics Committee met with developers and distributors of psychometric tests on 8 April 2005, at Johannesburg International Airport, following which the test publishers formed a group to place suggestions to the Board as part of the consultation process. Test publishers indicating at that stage that they wanted to be members are:

Prof HJ Kriek	SHL (Chairperson)
Dr L Fick	Corporate Solutions
Dr J De Beer	Jopie van Rooyen
Dr M de Beer	M&M Initiatives
Mr B Venter	PISA
Dr PF Erasmus	Potential Index Associates
Mr M Cunningham	Profiles International
Ms KE Dowdeswell	SHL
Prof SH van Deventer	Thomas International
Mr JC Schutte	Thomas International

ATP SA is currently in discussion with ATP International. There is already a European division and they are now busy with the creation of an African division. ATP International was established in 1992 in the USA. It is a non-profit organisation representing providers of tests and assessment tools and/or services related to education, employment, certification/licensing or clinical issues. The mission of the industrial division is:

*"Business organisations use a variety of tests as aids in hiring, placing or promoting employees. Industrial/Organisational test publishers are dedicated to promoting and advancing the role of quality assessment in the workplace."*

The mission statement of ATP International was accepted for ATP SA and was sent to the Board, indicating that the Test Publishers Association of South Africa accepts these principles and is in the process of negotiating a South African affiliation with ATP.

While the Industrial Division of the association was already been formed in 2005, the HPCSA requested that the association's membership be expanded to span all test publishers in South Africa. To address this request, and to define ATP SA's position concerning the reclassification of psychological tests, a meeting of test publishers was held on 19 June 2006 and all test publishers on the HPCSA list of distributors were invited to attend. The following test publishers attended:

Prof HJ Kriek	SHL (Chairperson)
Ms T Joubert	SHL
Dr L Fick	Corporate Solutions
Dr J De Beer	Jopie van Rooyen
Dr M de Beer	M&M Initiatives
Dr PF Erasmus	Potential Index Associates
Dr Renate Scherrer	Jopie van Rooyen
Ms Dawn Philip	Insights
Ms Theresa Cotterrell	MAC Assessment & Development
Dr N Claassen	Mindmuzik Media
Ms T Kriek	Mindmuzik Media
Dr M Prinsloo	Magellan Consulting
Dr W Barkhuizen	Saville Consulting
C Wilford	Profiles International
N Laycock	Amos Laycock Consulting
Mr H Bronkhorst	Organisation & Management Technology

Apologies were given by:

Ms N Tredoux	Psytech SA
Mr C Schutte	Thomas International
Dr T Taylor	Aprolab
Ms Laura Kartus	Placements Inc
Ms Lisa Ashton	Bioss

This document, laying out ATP SA's response to the request for feedback concerning the re-classification of tests, is based on the discussions held during the test publishers' meeting on 19 June 2006.

The following test publishers have read this final proposal to the Board and subscribe to the recommendations in the document:

Nic Laycock	Amos Laycock Consulting (Distributor)
Terry Taylor	Aprolab
Lisa Ashton	Bioss
Louis Fick	Corporate Solutions
Dawn Philip	Insights
Christina van Rooyen	Jopie van Rooyen
Kathy Pott	Jopie van Rooyen
Jopie de Beer	Jopie van Rooyen
Renate Scherrer	Jopie van Rooyen
Marie de Beer	M&M Initiatives
Theresa Cotterrell	MAC Assessment & Development
Maretha Prinsloo	Magellan Consulting
Talita Kriek	Mindmuzik Media
Frikkie Kriek	Mindmuzik Media
Nicolaas Claassen	Mindmuzik Media
Hendrik Bronkhorst	Organisation & Management Technology
Pieter Bronkhorst	Organisation & Management Technology
Steyn Bronkhorst	Organisation & Management Technology
Ben Venter	PISA
Laura Kartus	Placements Inc
Pieter Erasmus	Potential Index Associates
Colin Wilford	Profiles International
Mark Cunningham	Profiles International
Hennie Kriek	SHL
Tina Joubert	SHL
Carolyn Copeland	SOI
Margaret Bosshart	SOI
Vasi van Deventer	Thomas International
Chris Schutte	Thomas International
Lindiwe Msiza	Thomas International

The document has two sections, namely ATP SA's general view concerning test classification in South Africa, and a detailed commentary on the reclassification process proposal circulated by the Board.

## **General Standpoint of ATP SA Concerning Test Classification in South Africa**

- [1] A primary concern held by ATP SA is that the legal context in South Africa has been misunderstood and misinterpreted in the regulations published by the Board. This is evident in the classification and control of psychological tests as well as the use of psychological tests. Pertaining to this, ATP SA has requested Senior Counsel to formulate a legal opinion reflecting the correct interpretation of the relevant provisions. (Legal counsel will only be able to provide this on July 17 and this will then be forwarded to the Board).
- [2] Furthermore, there is a clear indication of a global norm to empower psychologists to take control of the total psychological testing process, not to legislate what is or is not allowed. There are countries that restrict use of tests to psychologists (e.g. Italy) but they do allow the psychologist to use their professional judgment regarding what are appropriate controls and risk mitigation strategies in assessment. In this document it is clear that South African regulation of Psychological tests is out of pace with what is happening in the rest of the world (See findings of a global study on psychological testing and regulations survey in 21 countries attached in Appendix A). Also see the document compiled by JvR exploring international trends in the grading of psychometric tests and qualification levels of test users, included here in Appendix B.

ATP SA feels that there should be a move towards self-regulation and self-classification, consistent with global trends and in particular with occupational tests. For example, in the USA while there are no statutory requirements or regulations regarding the use of psychological assessment, the Equal Employment Opportunity Commission, supported by the Equal Employment Opportunity Act, has jurisdiction over people

using assessments incorrectly or inappropriately in the employment context.

Relating to user competencies, ATP SA's opinion is that the psychologist must be empowered to be able to delegate responsibilities in psychological testing to individuals with the necessary knowledge, skills, experience etc required to do so, given the complexity and administration mode of the psychological test in question. This is in line with the medical practitioner that will delegate the injection of insulin to a diabetic patient.

- [3] ATP SA would like to recommend, following in line with international trends, that more psychological test categories are needed than the current 'psychological test' and 'non-psychological test'.

ATP SA feels strongly that a separate category is required for 'Industrial or Occupational Tests', which needs to receive special attention. It is felt that the public can be protected more efficiently under the auspices of the Employment Equity Act. The number of court cases on assessment, testing and selection decisions in industry is proof of this.

- [4] Finally, ATP SA recommends there should be more open communication of the usefulness of specific tests and classification of tests with the general public, as well as practitioners and psychologists. For example, the information provided by the European Federation of Psychological Associations' (EFPA) test review form provides valuable information for psychologists to use the classification information effectively (see Appendix C). This document is the conclusion of a long process that is currently subscribed to by all the EU countries. The United Kingdom also subscribes to this model and Denmark and Sweden are currently in the process of accepting this classification system. It is interesting to note that with the language and cultural diversity of the European Union all Psychological Federations were comfortable with this classification system.

## **ATP SA's Commentary on the Board's Reclassification Process Proposal**

Herewith follows ATP SA's detailed commentary on the reclassification process proposal circulated by the Board, entitled *'Re-classification of Psychological tests on the current list of classified tests (document for discussion and consultation)'* (dated 10 April 2006, Ref: 18/11/B).

In the following sections the general views of ATP SA regarding test classification are supported and expanded on, with specific reference to the Board's reclassification process proposal document. As such, some overlap with the general views of the test publishers may occur.

### ***Reasons for the need to re-classify psychological tests***

#### **Historically classified tests that have never been reviewed and evaluated**

The discontinuation of the A, B and C test categories of the Test Commission of South Africa led to tests in general being classified at the 'C' level, where the use is restricted to a psychologist or a psychometrist under the supervision of a psychologist. ATP SA recommends that the grading system (i.e. A, B and C, and even more categories if required) be reconsidered (See the last page of the European Federation of Psychological Associations' (EFPA) test review form that explains the required level of expertise to use the classified psychological test).

Tests that were classified by the Test Commission of South Africa (prior to 1996) were automatically included on the list of classified tests by the Psychometrics Committee, without going through the review and evaluation process that newly submitted tests have to undergo. While ATP SA agrees this needs to be reviewed, ATP SA would like to strongly recommend that the system of classifying tests should be reviewed in its entirety, before tests are re-evaluated and re-classified.

The association holds major concerns regarding the capacity of the HPCSA and Psychometrics Committee to classify new tests that are submitted for review while at the same time going back to review historically classified tests. The

typical time taken at this stage for classification is anything between 6 months and two years.

Over the past few years, since the scrapping of the A, B, C level of test categories, all tests in South Africa have been evaluated using a clinical and therapeutic psychometric model as a yardstick. However, there should be a clear difference between tests that are used in industry and tests that are clinical or diagnostic in nature. Group assessments, the use of actuarial selection models and the internet to collect personnel decision making information are all potential psychological assessments and it is not possible for the HPCSA to keep up with new test and assessment devices coming onto the South African market. Searching the World Wide Web on Google for the action words Personality Test indicated 5800000 websites in June 2005 and 16300000 in June 2006, a 500% increase. Personality tests of all qualities can already be accessed by the public, completed and feedback received without any control. ATP SA feels that tests used in industry – i.e. occupational tests – can be effectively evaluated and controlled by South African labour legislation. It is important to note that the Employment Equity Act (no. 55 of 1998) clearly states:

*“Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment being used*

- a) has been scientifically shown to be valid and reliable;*
- b) can be applied fairly to all employees; and*
- c) is not biased against any employee or group.”*

The effective use of test results is also effectively controlled by the different unfair discrimination clauses. While it is recognised that the role of the Board and the Psychometrics Committee is to protect the interest of the public, it is evident from this statement that the interest of the public is clearly protected by the Employment Equity Act (EEA). It is therefore felt that distinction should be made between clinical and industrial tests. Industrial tests where no clinical diagnosis is being made and where the use is for Human Resources interventions and decisions are covered in the EEA.

## **The need to accommodate newly created registration categories**

There exists confusion as to why accommodating the new professional registration categories requires the re-classification of tests. The current understanding is that everyone who is registered with the Professional Board of Psychology can obtain access to psychological tests (with or without supervision). ATP SA recommends revisiting the classification of users of different levels of tests, to move in line with global trends as well as to remove unnatural barriers to using objective assessment in industry effectively.

The roles, tasks and responsibilities of the different registration categories need to be clearly formulated and communicated to the public, practitioners, as well as test publishers. It should be noted that the tasks and responsibilities of the category Registered Psychological Counsellor are very vague, and poorly communicated to the public as well as the test publishers.

In a communication from the Board dated 8 June 1999 (Ref: 18/9/B Vol 2), the Board proposed a category of tester entitled *Test Administration Assistant*:

*“Test administration assistants will always work in tandem with psychologists and psychometrists and will be trained in and be responsible for generally assisting during the test administration process, giving test instructions, recording responses, and translating verbal responses and instructions where required.”*

The consultation draft of the European Test User Standards for test use in Work and Organizational settings, prepared by the European Federation of Psychologists' Associations (EFPA) and the European Association of Work and Organizational Psychologists, is based on the ITC's International Guidelines on Test Use. In this document, three standards are identified for following good practice in test administration, namely:

1. Make necessary preparations for the testing session
2. Administer the tests properly
3. Score the test results accurately

The document is designed to define what people need to know and do, what skills are required and what understanding is required for safe and competent use of a limited range of tests in a variety of occupational settings or contexts (More detail is provided in Appendix D – Competencies of Test Administrators).

ATP SA suggests that the whole issue of the psychologist being able to delegate to a test administrator or use the Internet to collect data and candidate responses be revisited to obtain clarity and overcome confusion in the market. In the United States the Task Force on Test User Qualifications suggests that a method of classification needs to be followed where the competencies required are defined in relation to the context, instrument, and use to which it will be put.

### **Ensuring consistency, clarity and legal compliance**

The motivation behind this reason for re-classification is unclear. No rights that have been awarded to certain people or organisations can be taken away through a change in process. A legal opinion in this regard has been requested and will be forwarded to the Board as soon as it's available (expected date of availability 17 July 2006).

With regard to the 'Occupational Test' category, ATP SA feels the whole system of classification needs to be revised. Based on this revision, stakeholders will be in a position to ensure classifications of tests are in line with legislation and current practice. As mentioned earlier, legal counsel obtained suggests that the Board's interpretation of the legislation is inaccurate. Once the system for classification has been revised, compliance will be achieved by empowering the psychologist to control the process.

An important point to note is that, due to the number and complexity of variables involved in people assessment in industry, it is virtually impossible to present the profession with specific, prescriptive guidelines on practice. An example of this in practice is presented by internet testing: establishing rules and regulations taking into account all complexities when testing across different platforms becomes unfeasible and impractical. ATP SA would like to suggest that the psychologist be empowered to ensure that practice conforms to fair and ethical test use and the professional should be ultimately accountable for their use of psychological tests.

## **Ensuring that tests are adequately supported and updated**

ATP SA agrees with the Board that the updating of tests is important, and that the test publishers have an important role to play. However, it is believed that the responsibility for updating tests and related support materials (i.e. test manuals and documentation) lies with the test publisher and not the Psychometrics Committee. As previously stated, the EEA demands responsibility from the test publisher to ensure their tests are valid, reliable and fair. Outdated tests that are not supported by up-to-date research cannot be viewed as valid, reliable and fair and will be exposed in the courts.

Again, ATP SA feels concern regarding the Psychometrics Committee's capacity to review tests that have already been classified while still keeping pace with reviewing tests newly submitted for classification. It is felt that it is more important to attempt to review all psychological tests that have not been classified in the past than to review psychological tests that have been submitted and classified already. ATP SA considers psychological tests currently being used that have never been classified to be more of a threat to public interest.

## ***Proposed process for the re-classification of tests***

### **Comment phase**

This document forms ATP SA's input to the Board's request for comments from all identified stakeholder groups.

### **Review and classification**

ATP SA holds concerns about the current test classification categories intended to be used in the re-classification process. We recommend that the attached "*EFPA Review Model for the Descriptions and Evaluation of Psychological Tests*" be used instead of the current form. ATP SA reviewed and evaluated various international processes and the EFPA model is a well-designed and thought through document that caters for various complexities, especially for South Africa's unique circumstances. Secondly, ATP SA strongly recommends that the difference

between occupational tests and clinical tests must be taken into consideration when looking at test classification (See Appendix C).

Additionally, ATP SA would like to recommend that the British Psychological Society (BPS) *Test Review Process* (2006) be followed, and a full-time independent review editor be appointed, instead of the current procedure. The BPS review process makes use of an Editorial Board that consists of a Senior Editor and a number of Consulting Editors (see Appendix E). The Psychometrics Committee has a history of not having the capacity to review and classify all the available tests. In this regard ATP SA would like to extend our assistance and a system of peer reviews by qualified individuals under a full time review editor could be considered.

It is very important that all documentation (except Intellectual Property) be made available to the Psychologists that use the instruments. Again, ATP SA would like to refer to the outstanding format as used by the EFPA. It is also recommended that the classification reviews with the star ratings as per the EFPA document be placed in the public domain to give access to this information to psychologists to help them to determine the possible use and suitability of classified tests.

### **Order of priority**

The order of priority should be as follows:

1. Tests that have never been reviewed or evaluated for classification
2. Historically classified tests prior to 1996
3. Tests with certificates that have been evaluated under the new system, after 1996, and identified as anomalous, ambiguous or in conflict with legislation or Board regulations.
4. Tests that have been classified under the current system (in the order that they were classified).

It is more important to review and classify tests that have never been classified than review tests that have already been through some classification process. The Psychometrics Committee does not have the capacity to review and classify new tests as well as those that have already been reviewed.

ATP SA feels strongly that the proposed new EFPA classification system be adopted and that the priorities as suggested be followed.

Also, ATP SA is not aware of any classification certificates that are in conflict with legislation (please see the legal opinion that will be forwarded to the Board).

## **Payment**

The process of re-evaluation must be in accordance with an undertaking from the Board with regards to confidentiality, time taken and costs involved. For both the evaluation process and costs, ATP SA requires the following:

*"In some instance, publishers may have technically important material that they are unwilling to make public for commercial reasons. In practice there is very little protection available for intellectual property to test developers (copyright law being about the only resource). Such reports might cover the development of particular scoring algorithms, test or item generation procedures and report generation technology. Where the content of such reports might be important in making a judgement in a review, the HPCSA and the full time classification editor plus reviewers should offer to enter into a non-disclosure agreement with the publisher. This agreement would be binding on the reviewers and editor. The reviewer could then evaluate the information and comment on the technological aspects and the overall evaluation to the effect that 'the starred rating in this review refers to materials held by publisher/distributor that have been examined by the reviewers on a commercial in confidence basis. These are not supplied with the manual."*

## Appendix A: Global Trends and the Regulation of Psychological Tests

Survey conducted by SHL, full report available on request.



"NewsLine\_Global trends.pdf"

# Global trends and the regulation of psychological tests

**In South Africa, the regulation of psychological tests is currently under discussion. To gain a broader perspective on the matter, SHL South Africa undertook to investigate the regulations for psychological testing in other countries. From the findings of the survey it may be concluded that South Africa's regulations are amongst the most stringent in the world.**

## Regulations in South Africa

The statutory body, the Health Professions Council of South Africa (HPCSA), was established to promote the health of the population, determine standards of professional education and training, and set and maintain standards of excellence for ethical and professional practice. The mandate of the Professional Board for Psychology is to control and exercise authority in respect of all matters relating to the training of psychologists, registered counsellors and psychometrists; promote the standards of such education and training in South Africa; and maintain and enhance the dignity of the profession and the integrity of the persons practising the profession.

## Survey methodology

To obtain a view of the regulations for psychological testing in other countries, a questionnaire was sent to country managers of SHL worldwide. The questionnaire consisted of ten questions and covered the topics of test regulation and classification, test administration, Internet testing, and feedback on test results.

Responses received were collated and, where applicable, the frequency of response types calculated. In total responses were received from 21 countries:

Australia	Belgium	Estonia
Finland	France	Greece
Hong Kong	India	Ireland
Israel	Italy	Japan
Netherlands	New Zealand	Portugal
Russia	Singapore	Sweden
Switzerland	UK	USA

## Statutory body or laws regulating the use of psychological tests

There are no statutory bodies or regulations concerning the use of psychological tests in ten of the countries that responded to the survey. In nine of the countries professional associations provide guidance for practitioners, which is not necessarily enforceable by law.

Regulations governing the use of psychological tests are provided by legislation in Israel and Finland; in Israel covering the use of psychological tests by psychologists. In Finland the law concerning privacy and data protection in the working life (2001) requires users of tests to be competent. However, the responsibility to ensure that test users are competent rests with the employer. The Finnish Psychologists Association, which is a member of the European Federation of Psychologists' Association, has established a certification board for the profession, but this is not yet well established.

## Classification system for psychological instruments

Fifteen of the countries that responded to the survey do not have a system that classifies psychological tests into different categories. Some indicated that while there are no formal classification systems, distinctions are made between personality and ability measures - Finland recommends distinguishing between competency-based and psychological (attribute-based) instruments.

Six of the countries responding have such a classification system; for example, Belgium's classification system differentiates between clinical personality questionnaires or projective tests,

occupational personality questionnaires, and personality questionnaires with report generation. In most cases where a system exists it also prescribes who may administer, score, interpret and provide feedback to candidates on psychological tests.

## Regulations for test administrators

Regulations or policies concerning who may administer psychological tests vary from country to country. However, nine of the countries that responded do not have any regulations or restrictions regarding who may administer psychological tests. In the USA there are no regulations for the administration of work-related assessments but in order to purchase clinical instruments such as the MMPI the purchaser is required to be a licensed psychologist. This is regulated by the test publishers.

In four of the countries responding to the survey the test publishers themselves are responsible for setting and monitoring the qualification requirements for use of their products.

Professional associations and psychologists' ethical codes give guidelines on who may administer psychological instruments in three of the countries included, although these guidelines are not necessarily enforceable by law. Five countries indicated the existence of some regulations concerning test administration.

## Qualifications required for a test administrator

The majority of respondents - fourteen countries - indicated that qualification requirements for test administrators were not specified. In two of the countries the test publishers set their own

requirements, while another two countries indicated that professional associations provide guidance. A further two countries require a degree in psychology and one country indicated that one had to be a qualified psychologist in order to administer psychological tests.

## Policies for online assessment

None of the countries surveyed had an official policy concerning Internet testing, nor one on controlled vs. supervised testing. SHL's corporate guidelines were named as providing guidance, as were the International Test Commission's International guidelines on computer-based and Internet delivered testing.

## Acceptance of online assessment

It is generally accepted by the HR community that the Internet is increasingly being used as a medium for recruitment and selection. In terms of the acceptance of online assessment, eleven of the countries responding indicated it was well received in their countries, while an additional country described it as enjoying a medium and growing acceptance. Online assessment is not widely used in seven of the responding countries.

Some of the reasons given for the acceptance and widespread use of online assessment included its ability to connect with markets and candidates, especially in geographically remote countries like New Zealand. India noted that online assessment is practically desirable. In Australia it is a popular approach for mass graduate recruitment drives. In the USA Internet testing is well accepted and growing in use, especially for professional jobs where the applicant pool uses the Internet extensively for job searching.

## Requirements for providing assessment feedback

The majority of respondents - ten of the countries responding - indicated that the requirements for providing assessment feedback to candidates were not specified in their country. Three of the countries participating in the survey require a

person to hold a degree in psychology, while Israel and Italy require the feedback provider to be a registered psychologist. In six countries, as with test administration, the test publishers set their own training requirements.

Test publishers make a distinction between who may use attribute and competency reports in terms of training received, but no legal or regulatory requirements are specified in any country.

## Conclusion

From the findings of the survey it has been concluded that South Africa's regulations regarding the use of psychological tests are amongst the most stringent in the world. It is interesting to note, however, the parallels between labour legislation in South Africa and the USA.

In the USA, while there are no statutory requirements or regulations regarding the use of psychological assessment, the Equal Employment Opportunity Commission (EEOC), supported by the Equal Employment Opportunity Act, has jurisdiction over people using assessments incorrectly or inappropriately in the employment context. The primary concern of the EEOC is to reduce non-job-related discrimination in hiring practices. This approach to testing provides greater control over the use of psychological assessment in the workplace as employers are forced by law to use best practices in their assessment methodology. This relates to the South African labour legislation context, where the Employment Equity Act prohibits the use of psychological testing or other similar assessments of an employee unless the test or assessment used can be shown to be valid and reliable, is applied fairly to all employees, and is not biased against any employee or group of employees. This parallel between the USA and South Africa's situation, which share the common goal of protecting employees from unfair discrimination, raises the question of whether relying on labour legislation to regulate psychological assessment results in greater control over the use of test level categories in the workplace.



## **Appendix B: Grading of Psychometric Tests and Qualification Levels of Test Users**

Compiled by JvR, work in progress.



"Grading of  
Psychometric tests J

**GRADING OF PSYCHOMETRIC TESTS AND  
QUALIFICATION LEVELS OF TEST USERS**

**Submitted to the HPCSA on 28 June 2006 by the  
Association of Test Publishers of South Africa**

**Research by JvR.**

# Table Of Contents

	Page
I) INTRODUCTION	2
II) PROFESSIONAL PSYCHOLOGICAL ASSOCIATIONS	2
A) HPCSA	2
B) AMERICAN PSYCHOLOGICAL ASSOCIATION (APA)	4
C) THE BRITISH PSYCHOLOGICAL SOCIETY	5
D) BRITISH COLUMBIA MINISTRY OF EDUCATION	6
E) THE AUSTRALIAN PSYCHOLOGICAL SOCIETY	8
F) NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH	9
G) KOSSOR EDUCATION NEWSLETTER	10
H) RECOMMENDATIONS FOR USERS OF PSYCHOLOGICAL TESTS – SARTA	10
III) ETHICAL STANDARDS IN TEST USE	11
A) NEW ZEALAND PSYCHOLOGICAL SOCIETY	11
B) INTERNATIONAL TEST COMMISSION	12
IV) CONCLUSION	12
V) REFERENCES	14

## I) INTRODUCTION

Grading or classification of psychological tests is used primarily as a means of safeguarding the public from those who may be inexperienced with the administration and interpretation of assessment tools or tests. The manner in which a psychological act is defined has an impact on whether the test is defined as being part of a psychological act or whether the test is defined as not psychological in nature. This in turn impacts on the grading and classification of a test. This type classification, while laudable in its attempt to restrict individuals who do not have the correct training from accessing tools, may also have the negative effect of restricting other professionals from accessing the very tests that are helpful in their contexts and have been developed specifically for their settings.

Currently in South Africa, the Health Professions Council and the Psychometrics Committee of the HPCSA regulate the tests that are available in the psychological arena.

Internationally, tests are graded or classified according to the levels as set down by the various psychological bodies as well as the test developers and distributors. The suppliers, publishers and developers of tests have certain grading levels they all use. Some use the same grading levels while others have their own levels. Most follow the alphabetized grading labels using, levels A, B, C etc. to demarcate which qualification levels are essential to have before a psychometric instrument can be purchased and used.

The information contained in this document has been sourced from the psychological associations of several countries and from test distributors worldwide.

## II) PROFESSIONAL PSYCHOLOGICAL ASSOCIATIONS

### A) HPCSA

The Health Professions Act, Act 56 of 1974, states that only registered psychologists are permitted to perform psychological acts which, in relation to evaluation, testing and assessment, are defined in Section 37 (2) (a), (b), (c), (d), and (e).

According to the Health Professions Act, tests, measures, questionnaires, instruments, etc. that tap psychological constructs must be used, interpreted, and controlled by psychologists. Certain psychological tests can, however, be used by psychometrists, psychotechnicians, and other professionals (e.g., speech and occupational therapists) provided that:

- (a) the use of the test has been certified for that category of tester by the Psychometrics Committee of the Professional Board for Psychology;
- (b) the tester complies with whatever restrictions may be placed on the test's use relevant to the category of test user that he/she is registered as. For example, a psychometrist may administer, score and preliminary interpret certain psychological tests but not report on the results of such tests;
- (c) the tester seeks mentoring from a psychologist where specialist input would enhance the testing process and the understanding of the test results; and
- (d) the tester has been appropriately trained and has achieved the minimum competencies required to use the test.

In view of the specific conditions under which psychological tests may be used by people other than registered psychologists, it is necessary to classify tests to facilitate the determination of the category of the tester who may use them. The Psychometrics Committee of the Professional Board of Psychology has been mandated by the Board to, among other things:

*'classify and advise on a regular revision of any device, instrument, method or test aimed at aiding persons or groups of persons in the adjustment of personality, emotional or behavioural problems or at the promotion of positive personality change, growth and development or for the determination of intellectual abilities, personality make-up, personality functioning, aptitude or interests.'* (HPCSA form 208, 2002, Pg 2, Par 2)

The Psychometrics Committee states "Classification of a test by the Psychometrics Committee of the Professional Board for Psychology does not impose any new restrictions on a psychological test (the Health Professions Act, Act 56 of 1974, imposes such restrictions). Instead, classification allows for the relaxing of conditions under which a test can be used, which makes the test more freely available."

There are two psychological test classification categories, namely:

1. Psychological Tests, which, although they can be used to varying extents by psychometrists and psychotechnicians (**See Addendum A Notes ii, iii and iv**), need to be under control of psychologists as regards:
  - a. selecting the test;
  - b. administering and scoring it;
  - c. interpreting it;
  - d. reporting on it.
2. Prescribed tests used by other Professionals. The Professional Board for Speech, Language and Hearing Professions, and Occupational Therapy and Medical Orthotics/Prosthetics, are in the process of furnishing the Psychometrics Committee with lists of tests that may tap psychological constructs which they use. The relevant Boards and the Psychometrics Committee will discuss these lists and reach agreement on the prescribed list of tests for the various professionals as well as the nature of the psychometrics and assessment training which trainees receive. (HPCSA Form 208, 2002, Pg 2, Par 4)

In terms of Qualification Levels that are necessary for access to the tests, Form 94 of the HPCSA Training and Examination guidelines states that for Psychometrists in the categories supervised practice and independent practice:

1. Psychometrists are not permitted to use certain personality measures (e.g., TAT, CAT, Rorschach); specialist neuropsychological measures; measures that are used for the diagnosis of psychopathology (e.g., MMPI-2)
2. Psychometrists are permitted to use (i.e., administer, score, and provisionally interpret) psychological tests under supervision and mentoring of a registered psychologist. When it comes to reporting the results, the psychologist needs to take final responsibility for the contents of the report, but the psychometrist may contribute to the content and co-sign the report. The reason why Psychometrists need to be mentored and supervised by psychologists is that the use of a psychological test constitutes a *psychological act*. (HPCSA Form 94, Pg 1, par 1)

## **B) AMERICAN PSYCHOLOGICAL ASSOCIATION (APA)**

### **American Psychological Association (APA) test User Qualifications Task Force – Report by: Dave Bartram, 21/01/01**

APA Ethical Principles of Psychologists and Code of Conduct:

The APA Task Force on test user qualifications defines the test user as:

“the person or persons responsible for the selection, administration and scoring of tests, for the analysis, interpretation, and communication of test results, and for any decisions or actions that are based, in part, on tests scores. Generally, individuals who simply administer tests, score tests and communicate simple or canned test results are not test users”. (APA Task Force report, Bartram, Updated 2005, Pg 1, Par 2)

The guidelines describe two areas of user competence: (a) generic competences that serve as a basis for most of the typical uses tests and (b) specific competences for the optimal use of tests in particular settings or for specific purposes.

The guidelines provide very detailed discussions of competence requirements for a number of different testing contexts (e.g. healthcare, counseling, employment). The competence needed by any particular test user will depend on the use that they will be making of tests and the context in which they will be doing the testing. They note that the testing process may be distributed between different individuals and make clear that the APA guidelines are directed primarily at the person who is responsible for the use of tests in the assessment process: Various activities included in the testing processes may be appropriately conducted by different people working collaboratively. Each participant should possess the knowledge, skills, and abilities relevant to his or her role.

The task force takes a position very similar to that adopted by the BPS in the UK, noting that evidence suggests that most of the problems associated with test use are related to the competence of individual tests users, rather than the quality of tests. However, it is also true that the uneven quality of test construction and the ease with which test instruments can be obtained from some test publishers contribute to these problems. (APA Task force Report, Bartram, updated 2005, Pg 1, Par 2)

The report by the task force discusses the three level system A, B, C, for classifying test user qualifications. These levels were first defined in 1950. According to the report, “This system labeled some tests (level A) as appropriate for administration and interpretation by non-psychologists; others (level B) as requiring some technical knowledge of test construction and use of supporting psychological and educational subjects such as statistics, individual differences, the psychology of adjustment, personnel psychology, and guidance; and others (level C) as being restricted to persons with at least a master’s degree in psychology, who have had at least one year of supervised experience under a psychologist. While the APA has dropped this method of classification from 1974 and subsequent editions of the Standards for Educational and psychological Testing, it has remained in use by many test publishers. The Task Force’s report suggests that this method of classification is now obsolete and needs to be replaced by one where the competences required are defined in relation to the context, instrument and use of which it will be put. (APA Task force report, Bartram, updated 2005, Pg 2, Par 2)

## **C) THE BRITISH PSYCHOLOGICAL SOCIETY**

### **The BPS Certificate of Competence – In Occupational Testing**

“The British Psychological Society Steering Committee on Test Standards has spent much of its time discussing the problem of how to ensure that those who use psychological tests do so to acceptable standards of competence”. (BPS, General Information Pack, Pg 5, Par 1)

Considerable concern was expressed by members of the Steering Committee, test publishers and many **Occupational Psychologists** about the growing problem of misuse and abuse of psychological testing in industry and commerce. These problems arose largely from the fact that purchasers of assessment materials and procedures often lacked the knowledge to make informed judgments about the worth of the products they were buying and the users of test materials were often inadequately trained – if trained at all – in their use and applications. (BPS, General Information Pack, Pg 5, Par 3)

The following certification scheme was developed from these initial discussions and focuses specifically on **occupational testing**.

#### **LEVEL A**

**LEVEL A** covers general foundations of testing, and the performance skills associated with test administration and interpretation of ability tests. A Checklist of Competences in Occupational Training (**LEVEL A**) are identified. The competences are those which are regarded as ‘essential’ and define the minimum levels of competence required for any test user. These cover the basic knowledge and skills which are necessary for all forms of psychological testing – whether paper-and-pencil or computer-based.

Any person who has met the standards required for all the items on the Checklist is eligible to apply to the Society for the British Psychological Society Certificate of Competence in Occupational Testing (**LEVEL A**). (BPS, General Information Pack-Level A, Pg 6, Par 1)

The Checklist refers to the following competences in terms of Psychological testing:

- defining the assessment needs;
- the basic principles of scaling and standardization;
- the importance of reliability and validity;
- deciding when psychological tests should or should not be used as part of an assessment process;
- administering tests to one or more candidates and dealing with scoring procedures;
- making appropriate use of test results and providing accurate written and oral feedback to clients and candidates; and
- maintaining security and confidentiality of test materials and test data. (BPS, General Information Pack-Level A, Pg 9, Par 1-7)

#### **LEVEL B**

**LEVEL B** has been designed to compliment **LEVEL A**. Together they provide a set of standards which define the competent user of psychological tests in occupational settings. While **LEVEL A** covers general foundations of testing and the performance skills associated with test administration and interpretation for group ability tests, **LEVEL B** is intended to increase the scope to cover personality assessment and the interpretation and use of personality tests. (BPS, General Information Pack-Level B, Pg 4, Par 1)

Three **LEVEL B** qualifications are available:

- an **Intermediate** level of qualification (Intermediate Level B) for people with sufficient foundation knowledge and expertise in the use of one appropriate instrument;
- an **Intermediate Plus** qualification for people with sufficient foundation knowledge and expertise in the use of a second or subsequent appropriate instrument;
- a **Full LEVEL B** qualification signifying a level of general competence significantly broader than that of the person who has been trained only in the use of specific personality assessment instruments. (BPS, General Information Pack-Level B , Pg 4, Par 4)

In Summary, **LEVEL A** covers the basic underpinning knowledge of psychometrics and practical skills associated with the use of tests (administration, scoring and analysis). The practical elements of **LEVEL A** focus on tests of 'maximum performance'. **LEVEL B** extends this to include tests of 'typical performance' – especially those requiring interpretation in terms of psychological theories or models.

**LEVEL B** is considered to incorporate and build on **LEVEL A**. It is not an alternative qualification: the **LEVEL B** standards do not therefore, repeat elements contained in **LEVEL A** except where matters covered at **LEVEL A** need to be considered in relation to issues associated with tests of 'typical performance', or where the basic issues of test choice and evaluation have to be faced. (BPS, General Information Pack, Pg 4, Par 8)

#### **D) BRITISH COLUMBIA MINISTRY OF EDUCATION**

**The Ministry Of Education in British Columbia's Manual of policies, procedures and Guidelines states the following in terms of the CLASSIFICATION OF EDUCATIONAL AND PSYCHOLOGICAL TESTS (LEVELS A – C)**

These guidelines relate to the use of psychological tests and the training of persons in testing and assessment practices. The guidelines focus on distinctions among levels of tests and their related restrictions; test user's qualifications and responsibilities; and restrictions for the purchase of tests. These guidelines are based on the American Association, 1985, and Canadian Psychological Association, 1987, standards and guidelines for educational and psychological testing." (BC Ministry of Education: Special Education Manual, Retrieved 2005, Pg 1 Par 1)

#### **LEVEL A TESTS**

These are tests that can be adequately administered, scored and interpreted with the aid of the manual, a familiarity with the client population, orientation to the kind of setting within which the testing is done, and a general knowledge of measurement principles and of the limitations of test interpretations. According to PRTC Procedures and Guidelines of August 2004; 'Such tests and aids may be given and interpreted by responsible non-psychologists (e.g., school principals, business executives).' This category includes most interest inventories, group or individual, and multiple-choice tests that employ a simple metric as the main avenue of interpretation (e.g., occupational clusters).

#### Administration of tests:

Training Standards – No training beyond advanced level course (senior undergraduate or graduate) in testing from an accredited college or university, or equivalent training under the direction of a qualified supervisor or consultant.

Interpretation of tests:

Training Standards – Minimum Standard: No training beyond advanced level course (senior undergraduate or graduate) in testing from an accredited college or university, or equivalent training under the direction of a qualified supervisor or consultant.

Best Practice: Minimum standard plus regular consultation for quality assurance with someone who has **LEVEL B or LEVEL C** training.

Purchaser eligibility:

Depends on publisher, but generally no restrictions except that the test must be ordered by the employee of the company directly responsible for administration of tests or an individual qualified to purchase **LEVEL B or LEVEL C** tests. (BC Ministry of Education: Special Education Manual, Retrieved 2005, Pg 1 Par 1-5)

## **LEVEL B TESTS**

These are tests which require specific training for administration, scoring and interpretation. These tests are more complex than **LEVEL A** tests and require sophisticated understanding of psychometric principles, the traits being measured, the client population and clinical issues involved in the setting within which the testing is done. This category would generally include most individual or group tests of achievement or interest, screening inventories and personnel tests.

Training standards:

Advanced level (senior undergraduate or graduate) course in testing from accredited college or university, or equivalent training under the direction of a qualified supervisor or consultant. Minimum preparation includes training in psychometric principles (reliability, validity, test construction), and direct supervised experience in administering, scoring and interpreting tests.

Interpretation of tests:

Training Standards – Minimum Standard: advanced level (senior undergraduate or graduate) course in testing from an accredited college or university, or equivalent training under the direction of a qualified supervisor or consultant including training in psychometric principles (reliability, validity, test construction), and direct supervised experience in administering, scoring and interpreting tests.

Best Practice: minimum standard plus regular consultation/supervision for quality assurance with someone with **LEVEL C** qualifications.

Purchaser eligibility:

These tests are available to individuals meeting the above training standards and to agencies where qualified test users are employed. (BC Ministry of Education: Special Education Manual, Retrieved 2005, Pg 1-2, Par 1-5)

## **LEVEL C TESTS**

Tests which require advanced (graduate level) training for interpretation in the specific professional field to which the tests apply (e.g., clinical psychology, counseling psychology, school psychology, industrial/organizational psychology). Some of these tests may also require this level of training for competent administration and scoring. According to the Psycho-educational Research and Training Centre (PRTC), Faculty of Education, University of British Columbia, Procedures and Guidelines of August 2004, *'Specifically, test publishers and professional standards indicate that use of LEVEL C tests for teaching or decision making purposes requires a minimum of a Masters and/or Doctoral (PhD) degree in education, psychology or related discipline. Also required is verification of licensure or registration as psychologist, or certification by a provincial school psychologists association or agency requiring training and experience in a relevant area of assessment consistent with the 1985 Standards.*

*The successful completion of a minimum of two university courses in test and measurement is critical.* (PRTC Procedures and Guidelines, Revised 2004, Pg 1, Par 4).

These tests are more complex than **LEVEL A AND B** tests. They require an in-depth understanding of psychometric principles, the traits and constructs being measured, the client population, and the clinical issues involved in the setting within which the testing is done. In addition, these tests require a high degree of professional skill and judgment for their interpretation. This group would generally include any aptitude or language or personality or clinical diagnostic test, group or individual.

#### Administration of the test

Training Standards – In situations where the person administering the test is closely supervised by a person qualified to use the test, some of these tests can be administered by someone with a minimum of a bachelor's degree in psychology or related discipline. Academic or supervised clinical experience must include:

- training in the theoretical constructs underlying the specific test instruments;
- training in the administration, scoring, and interpretation of these instruments;
- training in psychometric principles; and
- direct supervision of administration and (if appropriate) scoring.

#### Interpretation of tests

Training Standards – Minimum Standard: a minimum of a Master's degree in psychology or a related discipline and registration as a psychologist, or certification by the provincial school psychologists' association. Academic and supervised clinical experience must include:

- training in the theoretical constructs underlying the specific test instruments;
- training in the administration, scoring, and interpretation of these instruments;
- training in psychometric principles; and
- supervised administration, scoring and interpretation of these instruments.

#### Purchaser eligibility

These tests are restricted to individuals who meet the training standards above and who are members of qualified professional organizations and to agencies who have **LEVEL C** qualified test users overseeing the testing program. (BC Ministry of Education: Special Education Manual, Retrieved 2005, Pg 2 Par 5 –Pg 3 Par 2)

According to PRTC Procedures and Guidelines of August 2004; *'In some cases, however, LEVEL C tests can be administered by someone with a MA or BA degree or equivalent only IF they are closely supervised by a qualified consultant or supervisor.'* (PRTC Procedures and Guidelines, Revised 2004, Pg 1, Par 4)

## **E) THE AUSTRALIAN PSYCHOLOGICAL SOCIETY**

### **Classification of Psychological Tests**

According to the Australian Psychological association, "the level of skill and training required by psychologists for the administration, scoring, interpretation and reporting of tests varies according to the type of test, the method of administration and scoring, the nature of interpretation and the purposes for which the report is used. While the psychologist retains full responsibility for the interpretation and reporting of test scores, the administration and scoring of some of the tests may be delegated to suitably trained persons. The psychologist must be satisfied that any non-psychologist to whom delegation is made is aware of and abides by all relevant ethical and procedural considerations and ensure proper administration and scoring and maintenance of

confidentiality. Interpretation and reporting of the test results and the giving of feedback remain the responsibility of the psychologist. In general, delegation of test administration should be avoided; otherwise, all the principles of best practice in psychological testing apply. (Guidelines for the Use of Psychological Tests, Australian Psychological Society, 1997, Pg 2, Par 1.3)

## **F) NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH**

### **Supply of Tests**

In common with such organizations as the Australian Council for Educational Research and the National Foundation for Educational Research in England and Wales, the New Zealand Council for Educational Research supplies educational and psychological tests only to those persons who are qualified by training and experience to use and interpret them with professional care.

### **Test User Categories**

Different tests require varying amounts of training and experience. The New Zealand Council for educational research have therefore divided them into five broad categories:

**Level A:** Indicates tests which may be adequately administered, scored, and interpreted with the aid of a manual and with knowledge and experience of the field in which they are regarded as a useful aid. Completion of an introductory course in test administration and interpretation is considered a highly desirable prerequisite, e.g. Standardized test of achievement, standardized group of test of scholastic aptitude which report a single score, diagnostic test of achievement.

**Level B:** Indicates tests available only to those who have successfully completed a basic course in test administration and interpretation. Such a course would usually be a second or third year university course, requiring a supervised practical component and instruction in basic statistics essential; to test interpretation, e.g. group tests of intelligence which produce sub-scores, multiple-aptitude tests, test of specific aptitudes.

**Level Csp (Special note):** Some test from Level C may be available to individuals with basic B-level qualifications, provided that they have had special training in their use, e.g. speech, hearing, and language development (speech-language therapists) counseling inventories(guidance counselors) observational schedules and rating scales( teachers of children with special needs). Tests of this kind are classified as level Csp.

**Level C:** Indicates tests and measures available to users who have successfully completed advanced courses in the uses of tests and who have had, or are acquiring, casework or clinical experience, e.g. individual intelligence tests, attitude inventories, personality inventories, and scales.

**Level D:** Indicates clinical and diagnostic measures available only to those users who have undertaken special post graduate training in psychology or educational psychology, or equivalent on-the-job or in-service advanced training and experience, e.g., projective tests of personality. (NZCER, supply of Tests, Pg 1, Par 2)

## **G) KOSSOR EDUCATION NEWSLETTER**

**THE KOSSOR EDUCATION NEWSLETTER      Volume 3 Number 11**  
**Copyright 1995 S.A. Kossor   All Rights Reserved**

### **Using psychological materials in public schools**

Professional psychologists have known for more than 50 years that people can be harmed through the misuse or abuse of psychological tests. Access to psychological tests must be restricted to those who have been trained sufficiently in their use and potential for misuse. In an attempt to protect the public from harm, the American Psychological Association (APA) publishes Standards for Educational and Psychological Testing. Companies that sell psychological tests (The Psychological Corporation and Psychological Assessment Resources, for example) have established policy guidelines for the purchase of test materials, based on the APA Standards. Three levels of qualification for the purchase of psychological tests have been established:

Level A: No special qualifications required.

Level B: At least a B.A. degree in psychology, counseling or a closely related field, and relevant training or coursework in the interpretation of psychological tests and measurement at an accredited college or university; OR verification of membership in a professional association recognized to require training and experience in the ethical and competent use of psychological tests; OR licensed or certified by an agency which does the same.

Level C: A graduate degree in psychology, education, or closely related field that includes advanced training in the administration and interpretation of psychological tests; OR membership in a professional association that requires training and experience in the ethical and competent use of psychological tests, OR licensed or certified by an agency which does the same.

## **H) RECOMMENDATIONS FOR USERS OF PSYCHOLOGICAL TESTS – SARTA**

### Qualifications (Training) of Test Users

In general, the sophistication of the test will determine the level of qualification needed to administer. For complex tests, requiring extensive interpretation, considerable training will be needed. The American Counseling Association (ACA), the American Psychological Association (APA), and other groups have developed guidelines for the administration of psychological measures. It is prudent for anyone who utilizes psychological instruments to become familiar with standards set forth by these organizations. For further information on qualifications of test users, see Moreland et al. (1995).

### **III) ETHICAL STANDARDS IN TEST USE**

#### **A) NEW ZEALAND PSYCHOLOGICAL SOCIETY**

##### **Psychological Assessment**

Psychologists have the prime responsibility for psychological and behavioural assessment, including interviews, observations, standardized tests, questionnaires and psycho-physiological measures, and they ensure that these are used and interpreted only by competent persons.

Psychologists adequately protect the physical security of assessment instruments, the data they generate and psychological reports based on them.

Psychologists guard against any misuse or bias in selection, administration, scoring and interpretation of assessment instruments or procedures. They are prepared to justify, in terms of current scientific literature, their use and interpretation of any assessment instrument or procedure. They avoid using instruments which are obsolete or of dubious scientific status.

Psychologists obtain the informed consent of clients when undertaking psychological assessments. Informed consent means obtaining the consent of the client or, where the client is judged incapable of giving informed consent, the consent of those authorized to represent the interests of the client. Informed consent includes:

- a. Being informed of the nature and purpose of an assessment;
- b. Being informed of the procedures to be employed in the assessment process (e.g. type and general format of tests or questionnaires, psycho-physiological procedures etc.);
- c. Being informed of the uses to which the data from assessments will be put and the persons, organizations and/or agencies to whom the data and/or reports will be made available;
- d. Being informed of the right to know the content of psychological assessment reports concerning them.

In reporting assessment findings to clients and to other professionals, psychologists endeavor to ensure that appropriate explanations of the findings and their interpretations are provided and that they are not misused. Any reservations concerning the validity or reliability of an assessment procedure, arising from its administration, norms or domain-reference, should be made explicit in the report. Psychologists strive to prevent misuse of outdated assessment results.

Psychologists do not normally release uninterrupted data from assessments to persons who are not specifically trained in the use and interpretation of the instruments concerned.

Psychologists are responsible for ensuring adequate supervision of assessment instruments of procedures administered, scored or interpreted by others under their direction unless such persons are themselves properly trained in their use.

Psychologists abide by such guidelines for the standard of training required for the purchase, access to or use of assessment instruments as may be adopted from time to time by the Council.

## **B) INTERNATIONAL TEST COMMISSION**

### International Guidelines for Test Use

The ITC guidelines state the following in terms of their guidelines for test use:

- 1) The user take responsibility for ethical test use:
  - they should act in a professional and ethical manner;
  - ensure that they have the competence to use tests;
  - take responsibility for their use of tests;
  - ensure that tests materials are kept in securely; and
  - that the test results are treated confidentially.
- 2) The user follows good practice in the use of tests:
  - evaluate the potential utility of testing in an assessment situation;
  - choose technically sound tests appropriate for the situation;
  - give due consideration to issues of fairness in testing;
  - make necessary preparation for the testing session;
  - administer the test properly;
  - score and analyze test results accurately;
  - interpret results appropriately;
  - communicate the results clearly and accurately to relevant others; and
  - review the appropriateness of the test and its use. (ITC, international guidelines for test use, version 2000, Retrieved 2005)

## **IV) CONCLUSION**

The need for control is essential in the use of psychometric instruments. Most tests that are available internationally, are graded appropriately for the levels of expertise that is required to use the tests. This aims to ensure that potential misuse of tests is avoided. However it is imperative that fair and reasonable access to psychometric materials is made available to professional users who are qualified in the use of these materials. Rules that are overly restricted and unfairly restrict other qualified professionals (e.g. Human resource departments, teachers, etc.) from using assessments that have, in many instances, been developed especially for these settings, are likely to have an adverse effect rather than safeguarding individuals.

Furthermore, in the event of non-psychologists, for instance masters students or doctoral students that are not registered with the HPCSA, but who may have had years of experience in the psychological field, or may have developed some of the tools available, restriction of access to psychometric tests would not be logical.

## V) REFERENCES

1. HPCSA – THE PROFESSIONAL BOARD FOR PSYCHOLOGY – TRAINING AND EXAMINATION GUIDELINES FOR PSYCHOMETRISTS IN THE CATEGORIES SUPERVISED PRACTICE AND INDEPENDENT PRACTICE. – FORM 94, Retrieved 2005, from: [www.hpcsa.com](http://www.hpcsa.com)
2. HPCSA – THE HEALTH PROFESSIONS COUNCIL OF SOUTH AFRICA – POLICY ON THE CLASSIFICATION OF PSYCHOMETRIC MEASURING DEVICES, INSTRUMENTS, METHODS AND TECHNIQUES. – FORM 208, Update: November 2002, retrieved 2005, from: [www.hpcsa.com](http://www.hpcsa.com)
3. HPCSA – THE PROFESSIONAL BOARD OF PSYCHOLOGY – HEALTH PROFESSIONS COUNCIL OF SOUTH AFRICA – LIST OF TESTS CLASSIFIED AS BEING PSYCHOLOGICAL TESTS – FORM 207, Update: 22 November 2003, retrieved 2005, from: [www.hpcsa.com](http://www.hpcsa.com)
4. Harcourt Assessment, Inc, Terms and conditions of Sale, Retrieved 2005, From:<http://harcourtassessment.com>
5. PSYCHOLOGICAL TESTING: A USER'S GUIDE, retrieved 2005 from: [www.psychtesting.org.uk](http://www.psychtesting.org.uk)
6. GENERAL INFORMATION PACK – Certificate of Competence in Occupational Testing (LEVEL A AND LEVEL B), Retrieved 2005, from: [www.psychtesting.org.uk](http://www.psychtesting.org.uk)
7. PSYCHOEDUCATIONAL RESEARCH AND TRAINING CENTRE (PRTC) Faculty of Education, University of British Columbia. Retrieved 2005, From: [http://prtc.educ.ubc.ca/pdf\\_files/ethical\\_use.pdf](http://prtc.educ.ubc.ca/pdf_files/ethical_use.pdf)
8. SPECIAL EDUCATION SERVICES: A MANUAL OF POLICIES, PROCEDURES AND GUIDELINES Government of British Columbia, Retrieved 2005,from: [http://www.bced.gov.bc.ca/specialed/ppandg/x0012\\_12\\_appendix\\_b.htm](http://www.bced.gov.bc.ca/specialed/ppandg/x0012_12_appendix_b.htm)
9. Retrieved 2005, From: <http://www.acer.edu.au/publications/acerpress/qualifications.html>
10. Retrieved 2005, From: <http://www.apa.org/ethics/code2002.html>
11. Retrieved 2005, From: [http://www.psychology.org.nz/about/code\\_1986.html](http://www.psychology.org.nz/about/code_1986.html)
12. Retrieved 2005, From: [http://www.intestcom.org/APA\\_Test\\_Use\\_Task\\_Force.htm](http://www.intestcom.org/APA_Test_Use_Task_Force.htm)
13. Retrieved 2005, From: <http://www.insightassessment.co./restrictions.html>
14. Retrieved 2005, From: <http://www.psychometrics.co./privacy/index.cfm>
15. Retrieved 2005, From: <http://.myacpa.org/comm/assessment/dragon/dragon-recs.html>
16. Retrieved 2005, From: [http://www.voicenet.com/~sakossor/Pe3\\_11.html](http://www.voicenet.com/~sakossor/Pe3_11.html)
17. Retrieved 2005, From: <http://www.dos.state.pa.us/bpoa/LIB/bpoa/20/10/psybd-act.pdf>
18. Retrieved 2005, From: <http://www.pearsonassessments.com/forms/levels.htm>

19. Retrieved 2005, From: <http://www.sigmaassessmentsystems.com/departments/qualifications.asp>
20. Retrieved 2005, From: [http://www.pantesting.com/auth\\_users.htm](http://www.pantesting.com/auth_users.htm)

## **Appendix C: EFPA Review Model for the Description and Evaluation of Psychological Tests**

European Federation of Psychologists' Associations



"EFPA TEST REVIEW  
FORM and NOTES Ve

# EFPA REVIEW MODEL FOR THE DESCRIPTION AND EVALUATION OF PSYCHOLOGICAL TESTS<sup>1</sup> TEST REVIEW FORM AND NOTES FOR REVIEWERS<sup>2</sup>

It is **IMPORTANT** that reviewers read all notes carefully whilst carrying out the review.

Original version compiled and edited by Dave Bartram  
Updated and revised by Patricia Lindley, Dave Bartram and Natalie Kennedy, April  
2004, May 2005<sup>3</sup>  
Current version 3.41: August 2005

---

For details of the Review Procedure and any local modifications to the review content and criteria, consult your local Psychological Association. The present document provides for procedures that employ two reviewers for each test review, with a third person to oversee the review (the 'Consulting Editor') and a Senior Editor who is responsible for ensuring uniformity of application of the criteria across reviews. Local arrangements may result in some of these functions being combined.

EFPA recommends that the evaluations in these reviews are directed towards qualified practising test users, though they should also be of interest to academics and specialists in psychometrics and psychological testing.

Users of this document and its contents are required by EFPA to acknowledge this source with the following text:

*"The EFPA Test Review Criteria were largely modelled on the form and content of the British Psychological Society's (BPS) test review criteria and criteria developed by the Committee of Test Affairs (COTAN) of the Dutch Association of Psychologists (NIP). Dave Bartram and Patricia Lindley originally developed the BPS criteria and review procedures for the UK Employment Service and later expanded them for use by the BPS. Arne Evers edited the Dutch rating system for test quality. EFPA is grateful to the BPS and the NIP for permission to build on their criteria in developing the European model. EFPA is also grateful to Dave Bartram, Arne Evers, and Patricia Lindley for their contributions to the development of this model. All intellectual property rights in the original BPS and NIP criteria are acknowledged and remain with those bodies."*

---

<sup>1</sup> The EFPA Standing Committee on Tests and Testing has endorsed these notes for reviewers and the related review format. Member Psychological Associations may use them as the basis for their own instrument review procedures. The intention of making this widely available is to encourage the harmonisation of review procedures and criteria across Europe. Comments on these documents are welcomed in the hope that the experiences of instrument reviewers will be instrumental in improving and clarifying the processes.

<sup>2</sup> This document has been produced from a number of sources, including the BPS Test Review Evaluation Form (NPAL, and BPS Steering Committee on Test Standards); the Spanish Questionnaire for the Evaluation of Psychometric Tests (Spanish Psychological Association) and the Rating System for Test Quality (Committee on Testing of the Dutch Association of Psychologists). Some of the content has been adapted with permission from: BPS Books Reviews of Level B Assessment Instruments for use in Occupational Assessment, Notes for Reviewers: Version 3.1. December 1998: Copyright © NPAL, 1989, 1993, 1998.

<sup>3</sup> The present version is an integration of two separate documents (the Review Form and the Notes for Reviewers). In addition the contents have been edited and amended in the light of use by the BPS for their online test reviews. Version 3.4 was accepted as the replacement for all previous versions by the General Assembly of EFPA in July 2005.

.....

## Section 1: Description of the Instrument: General Information & Classification

.....

The first section of the form should provide the basic information needed to identify the instrument and where to obtain it. It should give the title of the instrument, the publisher and/or distributor, the author(s), the date of original publication and the date of the version that is being reviewed.

Sections 1.1 through 1.9 should be straightforward. They are factual information, although some judgement will be needed to complete information regarding content domains.

EFPA 3.2 reference

	<b>Reviewer 1:</b>	
	<b>Reviewer 2:</b>	
	<b>Consulting Editor:</b>	
	<b>Senior Editor:</b>	
	<b>Senior Update Editor:</b> (only required for updates)	
	<b>Update Editor:</b> (only required for updates)	
	<b>Date of Current Review:</b>	
1.1	<b>Instrument Name (local version):</b>	
	<b>Short Version Test Name:</b>	
1.2	<b>Original Test Name (if the local version is an adaptation):</b>	
1.4	<b>Authors of the Original Test:</b>	
1.3	<b>Authors of the Local Adaptation:</b>	
1.7	<b>Local Test Distributor/Publisher:</b>	
1.8	<b>Publisher of the original version of the Test (if different to current Distributor/Publisher):</b>	
1.9.1	<b>Date of publication of current revision/edition:</b>	
1.9.2	<b>Date of publication of adaptation for local use:</b>	
1.9.3	<b>Date of publication of original Test:</b>	

**General Description of the Instrument** Short stand-alone non-evaluative description (200-600 words)

This section should contain a concise non-evaluative description of the instrument. The description should provide the reader with a clear idea of what the instrument claims to be - what it contains, the scales it purports to measure etc. It should be as neutral as possible in tone. It should describe what the instrument is, the scales it measures, general points of interest or unusual features and any relevant historical background. This section may be quite short (c 200-300 words). However, for some of the more complex multi-scale instruments, it will need to be longer (c 300-600 words). It should be written so that it can stand alone as a description of the instrument. As a consequence it may repeat some of the more specific information provided in response to sections 1.1 - 1.29.

This item should be answered from information provided by the publisher and checked for accuracy.

## Section 2: Classification

1.10.1	<p><b>Contents Domain</b> <i>(please tick those that apply)</i></p> <p>You should identify the content domains specified by the publisher. Where these are not clear, this should be indicated and you should judge from the information provided in the manual (standardisation samples, applications, validation etc) what the most appropriate answers are for 1.10.1.</p>	<input type="checkbox"/> Scholastic attainment <input type="checkbox"/> General ability <input type="checkbox"/> Verbal ability <input type="checkbox"/> Numerical ability <input type="checkbox"/> Spatial ability <input type="checkbox"/> Non-verbal ability <input type="checkbox"/> Perceptual speed <input type="checkbox"/> Memory <input type="checkbox"/> Manual skills/dexterity <input type="checkbox"/> Personality – Trait <input type="checkbox"/> Personality – Type <input type="checkbox"/> Personality – State <input type="checkbox"/> Cognitive Styles <input type="checkbox"/> Motivation <input type="checkbox"/> Values <input type="checkbox"/> Interests <input type="checkbox"/> Beliefs <input type="checkbox"/> Disorder and pathology <input type="checkbox"/> Group function <input type="checkbox"/> Family function <input type="checkbox"/> Organisational function, aggregated measures, climate etc <input type="checkbox"/> School or educational function <input type="checkbox"/> Other: (describe below)
1.10.2	<p><b>Intended or Main Area (s) of Use.</b> <i>(please tick those that apply)</i></p> <p>You should identify the intended areas of uses specified by the publisher. Where these are not clear, this should be indicated and you should judge from the information provided in the manual (standardisation samples, applications, validation etc) what the most appropriate answers are for 1.10.2.</p>	<input type="checkbox"/> Psycho-clinical <input type="checkbox"/> Psycho-neurological <input type="checkbox"/> Forensic <input type="checkbox"/> Educational <input type="checkbox"/> Work and Occupational <input type="checkbox"/> Counselling, advice, guidance and career choice <input type="checkbox"/> General health, life and well-being <input type="checkbox"/> Sports and Leisure <input type="checkbox"/> Other (describe below)
1.10.3	<p><b>Intended Mode of Use (conditions under which the instrument was standardised and validated)</b> <i>(tick one box)</i></p> <p>This section is important as it identifies whether the instrument has been designed with the intention of it being used in unsupervised or uncontrolled administration conditions. This item should be answered from information provided by the publisher and checked for accuracy.</p>	<input type="checkbox"/> Unsupervised administration without control over the identity of the test taker and without full control over the conditions of administration (e.g. open access Internet delivered test; test available for purchase from bookstores)  <input type="checkbox"/> Controlled by unsupervised administration. Control over conditions (timing etc) and some control of identify of the test taker (e.g. tests administered over the Internet but only to known individuals - password restricted access)

		<input type="checkbox"/> Supervised and controlled administration. Test administration under the control of a qualified administrator or proctor  <input type="checkbox"/> Managed administration. Test administration only provided through specified testing centres (e.g. licensing and certification assessment programmes)
1.10.4	<b>Description of the populations for which the test is intended:</b> This item should be answered from information provided by the publisher. For some tests this may be very general (e.g. adults), for others it may be more specific (e.g. manual workers, or boys aged 10 to 14). This section should contain the stated populations. Where these may seem inappropriate, this should be commented on in the Evaluation section of the review.	
1.10.5	<b>Number of scales and brief description of the variable or variables measured by the instrument</b> This item should be answered from information provided by the publisher. Please indicate the number of scales (if more than one) and provide a brief description of each scale if its meaning is not clear from its name.. Reviews of the instrument should include discussion of other derived scores where these are commonly used with the instrument and are described in the standard documentation - e.g. 16PF criterion scores and Adjustment Specification Equation scores - but not scales which are 'add-ons' - e.g. the use of 16PF or OPQ scores to generate Belbin team-type measures.	
1.11	<b>Items format (select one)</b>  This item should be answered from information provided by the publisher. Note that it is important not to confuse multiple choice same scale with multiple choice different scale item formats. The latter, ipsative, formats require test takers to make choices between sets of two or more items drawn from <i>different</i> scales. For 'Forced-choice mixed-scale alternatives,' the test taker has to select which of two scales are more like them or with which of two statements they most agree. For multiple choice version, there may be three or more statements drawn from an equivalent number of different scales. Typically these statements may have to be ranked or the most- and least-like-me options selected.	<input type="checkbox"/> Open <input type="checkbox"/> Multiple choice, same scale alternatives <input type="checkbox"/> Bipolar adjectives <input type="checkbox"/> Likert ratings <input type="checkbox"/> Forced choice, mixed scale alternatives (ipsative) – see Notes for explanation. <input type="checkbox"/> Multiple choice, mixed scale alternatives (ipsative) – see Notes for explanation. <input type="checkbox"/> Adjective pair of sets, mixed scales (ipsative) <input type="checkbox"/> Other (describe below)
1.12	<b>No of Test Items:</b>  This item should be answered from information provided by the publisher. If the instrument has several scales, make clear whether you are indicating the total number of items or the number of items for each scale. Where items load on more than one scale, this should be documented.	

1.13	<p><b>Administration Mode(s):</b></p> <p>This item should be answered from information provided by the publisher. If any special pieces of equipment (other than those indicated in the list of options, e.g. VCR, tape recorder) are required, they should be described here. In addition, any special testing conditions should be described. 'Standard testing conditions' are assumed to be available for proctored/supervised assessment. These would include a quiet, well-lit and well-ventilated room with adequate desk-space and seating for the necessary administrator(s) and candidate(s).</p>	<input type="checkbox"/> Interactive Individual Administration <input type="checkbox"/> Supervised Group Administration <input type="checkbox"/> Computerised locally-installed application – supervised/proctored <input type="checkbox"/> Computerised web-based application – supervised/proctored <input type="checkbox"/> Computerised locally-installed application – unsupervised/self-assessment <input type="checkbox"/> Computerised web-based application – unsupervised/self-assessment <input type="checkbox"/> Other (indicate)
1.14	<p><b>Response Mode:</b></p> <p>This item should be answered from information provided by the publisher. If any special pieces of equipment (other than those indicated in the list of options, e.g. VCR, tape recorder) are required, they should be described here. In addition, any special testing conditions should be described. 'Standard testing conditions' are assumed to be available for proctored/supervised assessment. These would include a quiet, well-lit and well-ventilated room with adequate desk-space and seating for the necessary administrator(s) and candidate(s).</p>	<input type="checkbox"/> Oral Interview <input type="checkbox"/> Paper & Pencil <input type="checkbox"/> Manual operations <input type="checkbox"/> Computerised <input type="checkbox"/> Other (indicate)
1.15	<p><b>Time required for administering the instrument:</b></p> <p>This item should be answered from information provided by the publisher. The response to this item can be broken down into a number of components. In most cases, it will only be possible to provide general estimates of these rather than precise figures. The aim is to give the potential user a good idea of the time investment associated with using this instrument. Do NOT include the time needed to become familiar with the instrument itself. Assume the user is experienced and qualified. Preparation time (the time it takes the administrator to prepare and set out the materials for an assessment session).</p> <ul style="list-style-type: none"> <li>Administration time per session: this includes the time taken to complete all the items and an estimate of the time required to give instructions, work through example items and deal with any debriefing comments at the end of the session.</li> <li>Scoring: the time taken to obtain the raw-scores.</li> <li>Analysis: the time taken to carry out further work on the raw scores to derive other measures and to produce a reasonably comprehensive interpretation (assuming you are familiar with the instrument).</li> <li>Feedback: the time required to prepare and provide feedback to a candidate.</li> </ul> <p>It is recognised that time for the last two components could vary enormously - depending on the context in which the instrument is being used. However, some indication or comments will be helpful.</p>	Preparation: <input type="text"/> Administration: <input type="text"/> Scoring: <input type="text"/> Analysis: <input type="text"/> Feedback: <input type="text"/>

1.16	<p><b>Indicate whether different forms of the instrument are available</b> (genuine or pseudo-parallel forms, short versions, computerised versions, etc). If computerised versions do exist, describe briefly the software and hardware requirements:</p> <p>For Section 1.16, report whether or not there are alternative versions (forms) of the instrument available and describe the applicability of each form for different groups of people. In some cases, different forms of an instrument are meant to be equivalent to each other - i.e. alternate forms. In other cases, various forms may exist for quite different groups (e.g. a children's form and an adult's form). Where more than one form exists, indicate whether these are equivalent/alternate forms, or whether they are designed to serve different functions - e.g. short and long version; ipsative and normative version.</p> <p>Some instruments may be partly or fully computerised or available in computerised versions. For each of the four 'stages' of the assessment process, indicate the options available from the supplier. Note that CBTI packages, if available, should be indicated.</p>	
------	---	--

.....

### Section 3: Measurement & Scoring

.....

1.17	<p><b>Scoring procedure for the test:</b></p> <p>This item should be completed by reference to the publisher's information and the manuals and documentation.</p> <p>Bureau services are services provided by the supplier - or some agent of the supplier - for scoring and interpretation. In general these are optional services. If scoring and/or interpretation can be carried out ONLY through a bureau service, then this should be stated in the review - and the costs included in the recurrent costs section.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Computer scoring with direct entry of responses by test taker</li> <li><input type="checkbox"/> Computer scoring with manual entry of responses from the paper response form</li> <li><input type="checkbox"/> Computer scoring by Optical Mark Reader entry of responses from the paper response form</li> <li><input type="checkbox"/> Simple manual scoring key – clerical skills only required</li> <li><input type="checkbox"/> Complex manual scoring – requiring training in the scoring of the instrument</li> <li><input type="checkbox"/> Bureau-service – eg. Scoring by the company selling the instrument</li> <li><input type="checkbox"/> Other (describe below)</li> </ul>
1.18	<p><b>Scores:</b></p> <p>This item should be completed by reference to the publisher's information and the manuals and documentation.</p> <p>Brief description of the scoring system to obtain global and partial scores, correction for guessing, qualitative interpretation aids, etc).</p>	
1.19	<p><b>Score Transformation for standard scores:</b></p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Normalised – scores obtained by use of normalisation look-up table</li> <li><input type="checkbox"/> Not-normalised – scores obtained by linear transformation</li> </ul>

1.20	<p><b>Scales used</b> <i>(tick all that apply)</i></p> <p>This item should be completed by reference to the publisher's information and the manuals and documentation.</p>	<p><b>Percentile Based Scores</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Centiles</li> <li><input type="checkbox"/> 5-grade classification: 10:20:40:20:10 centile splits</li> <li><input type="checkbox"/> Deciles, and other equi-percentile classifications</li> </ul> <p><b>Standard Scores</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Z-scores</li> <li><input type="checkbox"/> IQ deviation quotients etc (e.g. mean 100, SD=15 for Weschler or 16 for Stanford-Binet)</li> <li><input type="checkbox"/> College Entrance Examination Board (e.g. SAT mean=500, SD=100)</li> <li><input type="checkbox"/> Stens,</li> <li><input type="checkbox"/> Stanines, C Scores</li> <li><input type="checkbox"/> T-Scores</li> <li><input type="checkbox"/> Other (please describe)</li> </ul>
------	--	--

---

## Section 4: Computer Generated Reports

---

**Note that this is purely *descriptive*. Evaluations of the reports would be part of the Evaluation section of the review**

For instances where there are multiple generated reports available please complete items 1.21 – 1.22 for each report (please copy pages if necessary). This classification system could be used to describe two reports provided by a system, for example, Report 1 may be intended for the test taker or other un-trained users, and Report 2 for a trained user who is competent in the use of the instrument and understands how to interpret it.

1.21	<p><b>Are Computer Generated Reports available with the instrument?</b></p> <p>If the answer to 1.21 is 'YES' then the following classification should be used to classify the types of reports available. For many instruments, there will be a range of reports available. Please complete a separate form for each report</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Yes (complete sections below)</li> <li><input type="checkbox"/> No (move to section 1.23)</li> </ul>
1.21.0	<p><b>Name or description of Report:</b> <i>(See introduction to this section)</i></p>	
1.21.1	<p><b>Media:</b> <i>(select one)</i></p> <p>Reports may consist wholly of text or contain text together with graphical or tabular representations of scores (e.g. sten profiles). Where both text and data are presented, these may simply be presented in parallel or may be linked, so that the relationship between text statements and scores is made explicit.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Text only</li> <li><input type="checkbox"/> Unrelated text and graphics</li> <li><input type="checkbox"/> Integrated text and graphics</li> <li><input type="checkbox"/> Graphics only</li> </ul>

1.21.2	<p><b>Complexity:</b><i>(select one)</i></p> <p>Some reports are very simple, for example just substituting a text unit for a sten score in a scale-by-scale description. Others are more complex, involving text units which relate to patterns or configurations of scale scores and which consider scale interaction effects.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Simple (For example, a list of paragraphs giving scale descriptions)</li> <li><input type="checkbox"/> Medium (A mixture of simple descriptions and some configural descriptions)</li> <li><input type="checkbox"/> Complex (Contains descriptions of patterns and configurations of scale scores, and scale interactions)</li> </ul>
1.21.3	<p><b>Report Structure:</b><i>(select one)</i></p> <p>Structure is related to complexity.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Scale based (where the report is built around the individual scales)</li> <li><input type="checkbox"/> Factor based (where the report is constructed around higher order factors - such as the 'Big Five' for personality measures.</li> <li><input type="checkbox"/> Construct based - where the report is built around one or more sets of constructs (e.g. in a work setting these could be such as team types, leadership styles, tolerance to stress etc) which are linked to the original scale scores.</li> <li><input type="checkbox"/> Criterion based where the reports focuses on links with empirical outcomes (e.g. training potential, job performance, absenteeism etc).</li> <li><input type="checkbox"/> Other</li> </ul>
1.21.4	<p><b>Sensitivity to Context</b> <i>(select one)</i></p> <p>When people write reports they tailor the language, form and content of the report to the person who will be reading it and take account of the purpose of the assessment and context in which it takes place. A report produced for selection purposes will be different from one written for guidance or development; a report for a middle-aged manager will differ from that written for a young person starting out on a training scheme and so on.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> One version for all contexts</li> <li><input type="checkbox"/> Pre-defined context-related versions</li> <li><input type="checkbox"/> User definable contexts and editable reports</li> </ul>
1.21.5	<p><b>Clinical-actuarial</b> <i>(select one)</i></p> <p>Most reports systems are based on clinical judgment. That is, one or more people who are 'expert-users' of the instrument in question will have written the text units. The reports will, therefore, embody their particular interpretations of the scales. Some systems include actuarial reports where the statements are based on empirical validation studies linking scale scores to, for example, job performance measures.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Based on clinical judgement of one expert</li> <li><input type="checkbox"/> Based on empirical/actuarial relationships</li> <li><input type="checkbox"/> Based on clinical judgement of group of experts</li> </ul>

1.21.6	<p><b>Modifiability</b> (<i>select one</i>)</p> <p>The report output is often fixed. However, some systems will produce output in the form of a file that can be processed by the user.</p>	<input type="checkbox"/> Not modifiable (fixed print-only output) <input type="checkbox"/> Limited modification (limited to certain areas e.g. biodata fields) <input type="checkbox"/> Unlimited modification (e.g. through access to Word Processor document file)
1.21.7	<p><b>Degree of Finish</b> (<i>select one</i>)</p> <p>A related issue is the extent to which the system is designed to generate integrated text - in the form of a ready-to-use report - or a set of 'notes', comments, hypotheses etc. The latter is clearly of far more use when the text is available to the user in modifiable form and can form the basis for the user's own report. In many cases, reports are designed to a very high standard of presentation, having a 'published' appearance and quality.</p>	<input type="checkbox"/> Publication quality <input type="checkbox"/> Draft Quality
1.21.8	<p><b>Transparency</b> (<i>select one</i>)</p> <p>Systems differ in their openness or transparency to the user. An open system is one where the link between a scale score and the text is clear and unambiguous. Such openness is only possible if both text and scores are presented and the links between them made explicit. Other systems operate as 'black boxes', making it difficult for the user to relate scales scores to text.</p>	<input type="checkbox"/> Clear linkage between constructs scores and text <input type="checkbox"/> Concealed link between constructs, scores and text <input type="checkbox"/> Mixture of clear/concealed linkage between constructs, scores and text
1.21.9	<p><b>Style and Tone</b> (<i>select one</i>)</p> <p>Systems also differ in the extent to which they offer the report reader guidance or direction. Some are stipulative: 'Mr X is very shy and will not make a good salesman...'. Others are designed to suggest hypotheses or raise questions: 'From his scores on scale Y, Mr X appears to be very shy. If this is the case, he could find it difficult working in a sales environment. This needs to be explored further with him.'</p>	<input type="checkbox"/> Directive <input type="checkbox"/> Guidance <input type="checkbox"/> Other
1.21.10	<p><b>Intended Recipients</b> (<i>select all that apply</i>)</p> <p>Reports are generally designed to address the needs of one or more categories of users. Users can be divided into four main groups:</p> <p>a) <i>Qualified users.</i> These are people who are sufficiently knowledgeable and skilled to be able to produce their own reports based on scale scores. They should be able to make use of reports that use technical psychometric terminology and make explicit linkages between scales and descriptions. They should also be able to customize and modify reports.</p> <p>b) <i>Qualified system users.</i> While not competent to generate their own reports from a set of scale scores, people in this group are competent to use the outputs generated by the system. The level of training required to attain this competence will vary considerably, depending on the nature of the computer reports (e.g. trait-based versus competency-based, simple or complex) and the uses to which its reports are to be put</p>	<input type="checkbox"/> Qualified test users <input type="checkbox"/> Qualified system users <input type="checkbox"/> Test takers <input type="checkbox"/> Third Parties

	<p>(low stakes or high stakes).</p> <p>c) <i>Test Takers</i>. The person who takes the instrument will generally have no prior knowledge of either the instrument or the type of report produced by the system. Reports for them will need to be in language that makes no assumptions about psychometric or instrument knowledge.</p> <p>d) <i>Third parties</i>. These include people - other than the candidate - who will be privy to the information presented in the report or who may receive a copy of the report. They may include potential employers, a person's manager or supervisor or the parent of a young person receiving careers advice. The level of language required for people in this category would be similar to that required for reports intended for Test Takers.</p>	
1.22	<b>Do Distributors offer a service to correct and/or develop computerised reports?</b> <i>(select one)</i>	<input type="checkbox"/> Yes <input type="checkbox"/> No

---

## Section 5: Supply, Condition and Costs

---

This defines what the publisher will provide, to whom, under what conditions and at what costs. It defines the conditions imposed by the supplier on who may or may not obtain the instrument materials. If one of the options does not fit the supply conditions, provide a description of the relevant conditions

1.23	<b>Documentation provided by the Distributor as part of the test package</b> <i>(select all that apply)</i>	<input type="checkbox"/> User Manual <input type="checkbox"/> Technical (psychometric) manual <input type="checkbox"/> Supplementary technical information and updates (eg. local norms, local validation studies etc) <input type="checkbox"/> Books and articles of related interest <input type="checkbox"/> Combinations of the above (specify)
1.24	<b>Methods of Publication</b> <i>(select all that apply)</i> For example, technical manuals may be kept up-to-date and available for downloading from the Internet, while user manuals are provided in paper form or on a CD.	<input type="checkbox"/> Paper <input type="checkbox"/> PC - Diskettes <input type="checkbox"/> PC – CD/ROM <input type="checkbox"/> Internet download <input type="checkbox"/> Live internet (instrument runs in a web browser) <input type="checkbox"/> Other (specify)
Sections 1.25- 1.27 cover costs. This is likely to be the section that is most quickly out of date. It is recommended that the supplier or publisher is contacted as near the time of publication of the review as possible, to provide current information for this section.		
1.25.1	<b>Start-Up Costs.</b> Price of a complete set of materials (all manuals and other material sufficient for at least one sample administration). Specify how many candidates could be assessed with the materials obtained for start-up costs, where these costs include materials for recurrent assessment  This item should try to identify the 'set-up' cost. That is the costs involved in obtaining a full reference set of materials, scoring keys and so on. It only includes training costs if the instrument is a 'closed' one - where there will be an <u>unavoidable</u> specific training cost, regardless of the prior	

	<p>qualification level of the user. In such cases, the training element in the cost should be made explicit. The initial costs do NOT include costs of general-purpose equipment (such as computers, cassette tape recorders and so on). However, the need for these should be mentioned. In general, define: any special training costs; costs of administrator's manual; technical manual(s); specimen or reference set of materials; initial software costs etc.</p>	
1.25.2	<p><b>Recurrent Costs:</b> Specify, where appropriate, recurrent costs of administration and scoring separately from costs of interpretation. (see 1.26 – 1.27)</p> <p>This item is concerned with the ongoing cost of using the instrument. It should give the cost of the instrument materials (answer sheets, non-reusable or reusable question booklets, profile sheets, computer usage release codes or 'dongle' units etc.) per person per administration. Note that in most cases, for paper-based administration such materials are not available singly but tend to be supplied in packs of 10, 25 or 50.</p> <p>Itemise any annual or per capita licence fees (including software release codes where relevant), costs of purchases or leasing re-usable materials, and per candidate costs of non-reusable materials.</p>	
1.26.1	<p><b>Prices for Reports generated by user installed software:</b></p>	
1.26.2	<p><b>Prices for Reports generated by postal/fax bureau service:</b></p>	
1.26.3	<p><b>Prices for Reports by Internet Service:</b></p>	
1.27	<p><b>Prices for other bureau services: correcting or developing automatic reports:</b></p>	
1.28	<p><b>Test-related qualifications required by the supplier of the test</b> <i>(select all that apply)</i></p> <p>1.28 concerns the user qualifications required by the supplier. For this section, where the publisher has provided user qualification information, this should be noted against the categories given. Where the qualification requirements are not clear this should be stated under 'Other' <i>not</i> under 'None'. 'None' means that there is an explicit statement regarding the lack of need for qualification.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> None</li> <li><input type="checkbox"/> Test specific accreditation</li> <li><input type="checkbox"/> Accreditation in general achievement testing: measures of maximum performance in attainment</li> <li><input type="checkbox"/> Accreditation in general ability and aptitude testing: measures of maximum performance in relation to potential for attainment</li> <li><input type="checkbox"/> Accreditation in general personality and assessment: measures of typical behaviour, attitudes and preferences</li> <li><input type="checkbox"/> Other (specify)</li> </ul>

1.29	<p><b>Professional qualifications required for use of the instrument</b> <i>(select all that apply).</i></p> <p>1.29 concerns the user qualifications required by the supplier. For this section, where the publisher has provided user qualification information, this should be noted against the categories given. Where the qualification requirements are not clear this should be stated under 'Other' <i>not</i> under 'None'. 'None' means that there is an explicit statement regarding the lack of need for qualification.</p>	<input type="checkbox"/> None <input type="checkbox"/> Practitioner psychologist with qualification in the relevant area of application <input type="checkbox"/> Practitioner psychologist <input type="checkbox"/> Research psychologist <input type="checkbox"/> Non-psychologist academic researcher <input type="checkbox"/> Practitioner in relevant related professions (therapy, medicine, counselling, education, human resources etc) <input type="checkbox"/> Holder of BPS Certificate of Competence in Occupational Testing Level A <input type="checkbox"/> Holder of BPS Certificate of Competence in Educational Testing Level A <input type="checkbox"/> Holder of BPS Certificate of Competence in Occupational Testing Level B <input type="checkbox"/> Other (indicate)
------	--	---

---

## Section 6: Evaluation of Test Materials

---

Potentially there are four sources of information that might be consulted in carrying out this evaluation:

1. The manual and /or reports that are supplied by the publisher for the user:
  - a) These are always supplied by the publisher /distributor before the instrument is accepted by the office and form the core materials for the review.
2. Open information that is available in the academic or other literature:
  - a) This is generally sourced by the reviewer and the reviewer may make use of this information in the review and the instrument may be evaluated as having (or having not) made reference to the information in its manual.
3. Reports held by the publisher that are not formally published or distributed:
  - a) The distributor/publisher may make these available at the outset or may send them when the review is sent back to the publisher to check for factual accuracy. The reviewer should make use of this information but note very clearly at the beginning of the comments on the technical information that "the starred rating in this review refers to materials held by the publisher/ distributor that is not [normally] supplied to test users". If these contain valuable information, the overall evaluation should recommend that the publisher should publish these reports and/or make them available to test purchasers
4. Reports that are commercial in confidence:
  - a) In some instances, publishers may have technically important material that they are unwilling to make public for commercial reasons. In practice there is very little protection available for intellectual property to test developers (copyright law being about the only recourse). Such reports might cover the development of particular scoring algorithms, test or item generation procedures and report generation technology. Where the content of such reports might be important in making a judgement in a review, the BPS could offer to undertake to enter into a non-disclosure agreement with publisher. This agreement would be binding on the reviewers and editor. The reviewer could then evaluate the information and comment on the technical aspects and the overall evaluation to the effect that "the starred rating in this review refers to materials held by the publisher/ distributor that have been examined by the reviewers on a commercial in confidence basis. These are not supplied with the manual."

### Explanation of Star Ratings

All sections are scored using the following rating system where indicated by: [rating]. Detailed descriptions giving anchor-points for each rating are provided.

Where a [ ] rating is provided on an attribute that is regarded as critical to the safe use of an instrument, the review will recommend that the instrument should not be used, except in exceptional circumstances by highly skilled experts or in research.

The instrument review needs to indicate which, given the nature of the instrument and its intended use, are the critical technical qualities. It is suggested that the convention to adopt is that ratings of these critical qualities are then shown in bold print.

In the following sections, overall ratings of the adequacy of information relating to validity, reliability and norms are shown, by default, in bold.

**Any instrument with one or more [ ] or [\*] ratings regarding attributes that are regarded as critical to the safe use of that instrument, shall not be deemed to have met the minimum standards.**

Entry on rating form	EPPA standard rating	UK Review representation	Explanation
[n/a]	[n/a ]	[n/a ]	This attribute is not applicable to this instrument
<b>0</b>	[ - ]	[None ]	Not possible to rate as no, or insufficient information provided
<b>1</b>	[ -1 ]	[* ]	Inadequate
		[** ]	NOT NOW USED
<b>3</b>	[ 0 ]	[*** ]	Adequate or Reasonable
<b>4</b>	[ 1 ]	[**** ]	Good
<b>5</b>	[ 2 ]	[***** ]	Excellent
		[N.r.i.o.r] * (for updates only)	Item was not rated in original review

In this section a number of ratings need to be given to various aspects or attributes of the documentation supplied with the instrument (or package). The term 'documentation' is taken to cover all those materials supplied or readily available to the qualified user: e.g. the administrator's manual; technical handbooks; booklets of norms; manual supplements; updates from publishers/suppliers and so on.

Suppliers are asked to provide a complete set of such materials for each Reviewer. If you think there is something which users are supplied with which is not contained in the information sent to you for review, please contact your Consulting Editor.

Items to be rated n/a or 0 to 5 (half ratings are acceptable)

### Rating

Quality of the explanation of the rationale, the presentation and the quality of information provided: (This overall rating is obtained by using judgment based on the ratings given for items 2.1 – 2.8)		Rating
2.1	<b>Overall rating of the Quality of the explanation of the rationale:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.1.1 – 2.1.5)	
2.1.1	i) Theoretical foundations of the constructs:	
2.1.2	ii) Test development procedure:	
2.1.3	iii) Thoroughness of the item analyses and item analysis model:	
2.1.4	iv) Explanation of content validity:	
2.1.5	v) Summary of relevant research:	

2.2	<p><b>Adequacy of documentation available to the user (user and technical manuals, norm supplements etc):</b> (This rating is obtained by using judgment based on the ratings given for items 2.2.1 – 2.2.6)</p> <p><i>For Section 2.2, the following 'benchmarks are provided for an 'excellent' (****) rating.</i> The focus here is on the quality of coverage provided in the documentation accessible to qualified users. Note that section 2.2. is about the comprehensiveness and clarity of the documentation available to the user (user and technical manuals, norm supplements etc.) in terms of its coverage and explanation. In terms of the quality of the instrument as evidenced by the documentation, areas in this section are elaborated on under: 2.1, 2.3, 2.9, 2.10 and 2.11.</p>	
2.2.1	<p><b>Rationale:</b> [see rating 2.1] Well-argued and clearly presented description of what is designed to measure and why it was constructed as it was.</p>	
2.2.2	<p><b>Development:</b> Full details of item sources, piloting, item analyses, comparison studies and changes made during development trials.</p>	
2.2.3	<p><b>Standardisation:</b> Clear and detailed information provided about sizes and sources of standardisation sample and standardisation procedure.</p>	
2.2.4	<p><b>Norms:</b> Clear and detailed information provided about sizes and sources of norms groups, conditions of assessment etc.</p>	
2.2.5	<p><b>Reliability:</b> Good explanation of reliability and a comprehensive range of internal consistency and retest measures provided with explanations of their relevance, and the generalisability of the assessment instrument</p>	
2.2.6	<p><b>Validity:</b> Good explanation of validity with a wide range of studies clearly and fairly described.</p>	
2.3	<p><b>Quality of the Procedural instructions provided for the user:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.3.1 – 2.3.7)</p>	
2.3.1	<p><b>For test administration:</b> Clear and detailed explanations and step-by-step procedural guides provided, with good detailed advice on dealing with candidates' questions and problem situations.</p>	
2.3.2	<p><b>For test scoring, norming etc:</b> Clear and detailed information provided, with checks described to deal with possible errors in scoring</p>	
2.3.3	<p><b>For interpretation and reporting:</b> Detailed advice on interpreting different scores, understanding normative measures and dealing with relationships between different scales, with plenty of illustrative examples and case studies</p>	
2.3.4	<p><b>For providing feedback and debriefing test takers and others:</b> Detailed advice on how to present feedback to candidates</p>	
2.3.5	<p><b>For providing good practice issues on fairness and bias:</b> <i>Detailed information reported about sex and ethnic bias studies, with relevant warnings about use and generalisation of validities</i></p>	
2.3.6	<p><b>Restrictions on use:</b> Clear descriptions of who should and who should not be assessed, with well-explained justifications for restrictions (e.g. types of disability, literacy levels required etc)</p>	
2.3.7	<p><b>References and supporting materials:</b> Detailed references to the relevant supporting academic literature and cross-references to other related assessment instrument materials.</p>	
<p><b>Quality of the materials:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.4 – 2.8)</p>		
2.4	<p><b>General quality of test materials</b> (test booklets, answer sheets, test objects, software, etc):</p>	
2.5	<p><b>Test quality of the local adaptation</b> (if the test has been translated and adapted into the local language):</p>	
2.6	<p><b>Ease with which the test taker can understand the task:</b></p>	
2.7	<p><b>Ease with which responses or answers can be made by the test taker:</b></p>	
2.8	<p><b>Quality of the items:</b></p>	

**Reviewers' comments on the documentation:** (comment on rationale, design, test development and acceptability)

---

---

## Section 7: Evaluation of Norms, Reliability and Validity

---

---

### General guidance on assigning ratings for these sections

It is almost impossible to set clear criteria for rating the technical qualities of an instrument. Under some conditions a reliability of 0.70 is fine; under others it would be inadequate. A criterion-related validity of 0.20 can have considerable utility in some situations, while one of 0.40 might be of little value in others. For these reasons, summary ratings should be based on your judgement and expertise as a reviewer and not simply derived by averaging sets of ratings.

These notes provide some guidance on the sorts of values to associate with inadequate, adequate, good and excellent ratings. However these are intended to act as guides only. The nature of the instrument, its area of application, the quality of the data on which reliability and validity estimates are based, and the types of decisions that it will be used for should all affect the way in which ratings are awarded.

In order to provide some idea of the range and distribution of values associated with the various scales that make up an instrument, enter the *number of scales* in each section. For example, if an instrument being used for group-level decisions had 15 scales of which 5 had retest reliabilities lower than 0.6, 6 between 0.6 and 0.70 and the other 4 in the 0.70 to 0.80 range, this would be entered as:

*Stability:*

*Median stability:*

[ ]	No information given.
[5]	Inadequate (e.g. $r < 0.6$ ).
[6]	Adequate (e.g. $0.6 < r < 0.70$ ).
[4]	Good (e.g. $0.70 < r < 0.80$ ).
[ ]	Excellent (e.g. $r > 0.80$ ).

It is realised that it may be impossible to calculate actual median figures in many cases. What is required is your best estimate, given the information provided in the documentation. There is space to add comments. You can note here any concerns you have about the accuracy of your estimates. For example, in some cases, a very high level of internal consistency might be commented on as indicating a 'bloated specific'.

For each of the possible ratings example values are given *for guidance only* - especially the distinctions between 'Adequate', 'Good' and 'Excellent'. It is recognised that what is a 'good' value for one instrument may be unrealistic for another. (For example, other things being equal we would expect a higher internal consistency for long scales than for short ones).

Where NORMS are concerned, the guidelines for sample sizes need to take into account the type of norms being used. If they claim to be representative general population norms, then the sample size will need to be quite large even to be 'adequate'. If they are occupation-specific norm groups, smaller sample sizes may be 'adequate'.

Careful consideration needs to be given to the suitability of international (same language) norms. Where these have been carefully established from samples drawn from a group of countries, they should be rated on the same basis as nationally based (single language) norm group.

For most purposes, samples of less than 150 will be too small, as the resolution provided in the tails of the distribution will be very small. The  $SE_{\text{mean}}$  for a z-score with  $n=150$  is 0.082 of the SD - or just better than one T-score point.

For VALIDITY, guidelines on sample sizes are based on power analysis of the sample sizes needed to find moderate sized validities if they exist. Concurrent and predictive validity refer to studies where real-world criterion measures (i.e. not other instrument scores) have been correlated with scales. Predictive studies generally refer to situations where assessment was carried out at a 'qualitatively' different point in time to the criterion measurement - e.g. for a work-related selection measure intended to predict job success, the instrument would have been carried out at the time of selection - rather than just being a matter of how long the time interval was between instrument and criterion measurement.

Construct validity includes correlations of scales from similar instruments. The guidelines on construct validity coefficients need to be interpreted flexibly. Where two very similar instruments have been correlated (with data obtained concurrently) we would expect to find correlations of 0.60 or more for 'adequate'. Where the instruments are

less similar, or administration sessions are separated by some time interval, lower values may be adequate. When evaluating construct validity care should be taken in interpreting very high correlations. Where correlations are above 0.90, the likelihood is that the scales in question are measuring exactly the same thing. This is not a problem if the scales in question represent a new scale and an established marker. It would be a problem though, if the scale in question were meant to be adding useful variance to what other scale already measure.

When judging overall validity, it is important to bear in mind the importance placed on construct validity as the best indicator of whether a test measures what it claims to measure. In some cases, the main evidence of this could be in the form of criterion-related studies. Such a test might have an 'adequate' or better rating for criterion-related validity and a less than adequate one for construct validity. In general, if the evidence of criterion-related validity or the evidence for construct validity is at least adequate, then, by implication, the overall rating must also be at least adequate. It should not be regarded as an average or as the lowest common denominator.

For RELIABILITY, the guidelines are based on the need to have a small Standard Error for estimates of reliability. Guideline criteria for reliability are given in relation to two distinct contexts: the use of instruments to make decisions about groups of people (e.g. classification of people into categories) and their use for making individual assessments. Reliability requirements are higher for the latter than the former. Other factors can also affect reliability requirements, such as whether scales are interpreted on their own, or aggregated with other scales into a composite scale. In the latter case the reliability of the composite should be the focus for rating not the reliabilities of the components.

Items to be rated n/a or 0 to 5 (half ratings are acceptable)

**Rating**

<p><b>Evaluation of Technical Information - Overall Adequacy:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.9 – 2.11)</p> <p><i>It is best to complete this rating after you have completed sections 2.9-2.11</i></p> <p>This section is concerned with the nature and quality of the technical information that is presented in the available documentation. It is not concerned with how clearly or how well the information is presented. <i>Please make sure that you base your ratings on the information readily obtainable by the general user of the assessment instrument rather than an academic or specialist.</i> Where other information is available about an instrument's qualities (e.g. American manuals not published in this country, articles in the psychological literature) this should be referred to in the <i>EVALUATION</i> section of the review.</p>	
---	--

## Norms or reference group information

2.9	<p><b>Overall adequacy:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.9.1 – 2.9.5 do not simply average numbers to obtain an overall rating.)</p> <p>Ratings can be defined, by entering number of scales that meet the following criteria, and then judging the rating from the mode of the distribution.</p> <p>Where an instrument is designed for use with out recourse to norms or reference groups, the 'not applicable' category should be used rather than 'no information given'</p>	[ ]
2.9.1	<p><b>Appropriateness for local use, whether local or international norms:</b></p> <p>[n/a] Not applicable  [ ] No information given.  [* ] Not locally relevant (e.g. inappropriate foreign samples).  [***] Local general population sample or non-local norms that could be used with caution.  [****] Local country samples or relevant international samples with good relevance for intended application.  [*****] Local country samples or relevant international samples drawn from well-defined samples from the relevant application domain.</p>	[ ]
2.9.2	<p><b>Appropriateness for intended applications:</b></p> <p>[n/a] Not applicable  [ ] No information given.  [* ] Norm or norms not adequate for intended applications.  [***] Adequate general population norms and/or range of norm tables.  [****] Good range of norm tables.  [*****] Excellent range of sample relevant, age-related and sex-related norms with information about other differences within groups (e.g. ethnic group mix).</p>	[ ]
2.9.3	<p><b>Sample sizes:</b></p> <p>[n/a] Not applicable  [ ] No information given.  [* ] Inadequate samples (e.g. less than 150).  [***] Adequate samples (e.g. 150-300).  [****] Large samples (e.g. 300-1000).  [*****] Very large samples (e.g. 1000+).</p>	[ ]
2.9.4	<p><b>Procedures used in sample selection: (select one)</b></p> <p>- No information is supplied  - Representative of population [summarise criteria below]</p> <p>- Incidental  - Random</p>	
2.9.5	<p><b>Quality of information provided about minority/protected group differences, effects of age, gender etc:</b></p> <p>[n/a] Not applicable  [ ] No information given.  [* ] Inadequate information.  [***] Adequate general information, with minimal analysis.  [****] Good descriptions and analyses of groups and differences  [*****] Excellent range of analyses and discussion of relevant issues relating to use and interpretation.</p>	[ ]

2.9.6 **Reviewers' comments on the norms:** Brief report about the norms and their history, including information on provisions made by the publisher/author for updating norms on a regular basis

## Validity

2.10	<b>Overall Adequacy:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.10.1 – 2.10.2.4. <b>Do not simply average numbers to obtain an overall rating. Usually this will be equal to either the Construct Validity or the Criterion-Related validity, whichever is the greater)</b>	
2.10.1	<b>Construct Validity - Overall Adequacy</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.10.1.2 – 2.10.1.6. <b>Do not simply average numbers to obtain an overall rating.</b> )	[ ]
2.10.1.1	<b>Designs used:</b> ( <i>tick as many as are applicable</i> ) <ul style="list-style-type: none"> <li>- No information is supplied</li> <li>- Correlations with other instruments and performance criteria</li> <li>- Intra-scale (item-rest correlations)</li> <li>- Differences between groups</li> <li>- Matrix Multitrait-Multmethod</li> <li>- Exploratory Factor Analysis</li> <li>- Confirmatory Factor Analysis</li> <li>- Experimental Designs</li> <li>- Other (indicate)</li> </ul>	
2.10.1.2	<b>Sample sizes:</b> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] One inadequate study (e.g. sample size less than 100).</li> <li>[***] One adequate study (e.g. sample size of 100-200).</li> <li>[****] More than one adequate or large sized study.</li> <li>[*****] Good range of adequate to large studies.</li> </ul>	[ ]
2.10.1.3	<b>Procedure of sample selection:</b> ( <i>select one</i> ) <ul style="list-style-type: none"> <li>- No information is supplied</li> <li>- Representative of population [summarise criteria below]</li>   <li>- Incidental</li> <li>- Random</li> </ul>	
2.10.1.4	<b>Median and range of the correlations between the test and other similar tests:</b> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] Inadequate (<math>r &lt; 0.55</math>).</li> <li>[***] Adequate (<math>0.55 &lt; r &lt; 0.65</math>).</li> <li>[****] Good (<math>0.65 &lt; r &lt; 0.75</math>).</li> <li>[*****] Excellent (<math>r &gt; 0.75</math>).</li> </ul>	[ ]
2.10.1.5	<b>Quality of instruments as criteria or markers:</b> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] Inadequate information given.</li> <li>[***] Adequate quality</li> <li>[****] Good quality.</li> <li>[*****] Excellent quality with wide range of relevant markers for convergent and divergent validation.</li> </ul>	[ ]
2.10.1.6	<b>Differential Item Functioning (DIF) analyses:</b> <ul style="list-style-type: none"> <li>[N/A ] Not applicable</li> </ul>	[ ]
2.10.2	<b>Criterion-related Validity - Overall Adequacy</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.11.1 – 2.10.4.2. <b>Do not simply average numbers to obtain an overall rating.</b> )	[ ]
2.10.2.1	<b>Description of the criteria used and characteristics of the populations:</b> ( <i>tick as many as are applicable</i> ) <ul style="list-style-type: none"> <li>- Concurrent</li> <li>- Predictive</li> <li>- Post-dictive</li> </ul>	

2.10.2.2	<p><b>Sample sizes:</b></p> <p>[ ] No information given.  [* ] One inadequate study (e.g. sample size less than 100).  [***] One adequate study (e.g. sample size of 100-200).  [****] One large or more than one adequate sized study.  [*****] Good range of adequate to large studies.</p>	[ ]
2.10.2.3	<p><b>Procedure of Sample selection:</b> (<i>select one</i>)</p> <ul style="list-style-type: none"> <li>- No information is supplied</li> <li>- Purposive or representative [summarise criteria below]</li>   <li>- Incidental</li> <li>- Random</li> </ul>	
2.10.2.4	<p><b>Median and range of the correlations between the test and criteria:</b></p> <p>[ ] No information given.  [* ] Inadequate (<math>r &lt; 0.2</math>).  [***] Adequate (<math>0.2 &lt; r &lt; 0.35</math>).  [****] Good (<math>0.35 &lt; r &lt; 0.50</math>).  [*****] Excellent (<math>r &gt; 0.50</math>)</p>	[ ]
2.10.3 <b>Reviewers' comments on validity:</b>		

## Reliability

2.11	<p><b>Overall Adequacy:</b>  <i>(This overall rating is obtained by using judgment based on the ratings given for items 2.11.1 – 2.10.3.2. Do not simply average numbers to obtain an overall rating.)</i>            For some instruments, internal consistency may be inappropriate (broad traits or scale aggregates), in which case place more emphasis on the retest data. In other cases (state measures), retest reliabilities would be misleading, so emphasis would be place on internal consistencies.</p> <p>In relation to reliability criteria, two main types of application are considered. Instrument that are designed for individual assessment require higher levels of reliability for practical effectiveness than those used to make decision on groups of people. In the suggested values given below, the first relates to instruments intended for making group decisions (e.g. selection sift tools) while the second set of values relates to those intended for scale-by-scale individual assessment.</p>	[ ]
2.11.1.	<p><b>Data provided about reliability:</b> <i>(select one)</i></p> <ul style="list-style-type: none"> <li>- Only one reliability coefficient given</li> <li>- Only one estimate of standard error of measurement given</li> <li>- Reliability coefficients for a number of different groups</li> <li>- Standard error of measurement given for a number of different groups</li> </ul>	
2.11.1	<p><b>Internal consistency:</b></p>	
2.11.1.1	<p><b>Sample size:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] One inadequate study (e.g. sample size less than 100).</li> <li>[**] One adequate study (e.g. sample size of 100-200).</li> <li>[***] One large or more than one adequate sized study.</li> <li>[****] Good range of adequate to large studies.</li> </ul>	[ ]
2.11.1.2	<p><b>Median of coefficients:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] Inadequate (e.g. <math>r &lt; 0.7</math>)</li> <li>[**] Adequate (e.g. <math>r = 0.7</math> to <math>0.79</math>)</li> <li>[***] Good (e.g. <math>r = 0.8</math> to <math>0.89</math>)</li> <li>[****] Excellent (e.g. <math>r &gt; 0.9</math>)</li> </ul>	[ ]
2.11.2	<p><b>Test retest stability:</b></p>	
2.11.2.1	<p><b>Sample size:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] One inadequate study (e.g. sample size less than 100).</li> <li>[**] One adequate study (e.g. sample size of 100-200).</li> <li>[***] One large or more than one adequate sized study.</li> <li>[****] Good range of adequate to large studies.</li> </ul>	[ ]
2.11.2.2	<p><b>Median of coefficients:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] Inadequate (e.g. <math>r &lt; 0.6</math>)</li> <li>[**] Adequate (e.g. <math>r = 0.6</math> to <math>0.69</math>)</li> <li>[***] Good (e.g. <math>r = 0.7</math> to <math>0.79</math>)</li> <li>[****] Excellent (e.g. <math>r &gt; 0.8</math>)</li> </ul>	[ ]
2.11.3	<p><b>Equivalence reliability:</b></p>	
2.11.3.1	<p><b>Sample size:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] One inadequate study (e.g. sample size less than 100).</li> <li>[**] One adequate study (e.g. sample size of 100-200).</li> <li>[***] One large or more than one adequate sized study.</li> <li>[****] Good range of adequate to large studies.</li> <li>[N/A] Not applicable</li> </ul>	[ ]
2.11.3.2	<p><b>Median of coefficients:</b></p> <ul style="list-style-type: none"> <li>[ ] No information given.</li> <li>[* ] Inadequate (e.g. <math>r &lt; 0.6</math>)</li> <li>[**] Adequate (e.g. <math>r = 0.6</math> to <math>0.69</math>)</li> <li>[***] Good (e.g. <math>r = 0.7</math> to <math>0.79</math>)</li> <li>[****] Excellent (e.g. <math>r &gt; 0.8</math>)</li> <li>[N/A] Not applicable</li> </ul>	

**Reviewers' comments on Reliability:**

- Comment on confidence intervals for reliability co-efficient
- Provide Spearman-Brown equivalents

---



---

**2.12 Section 7: Quality of Computer Generated Reports:**


---



---

For each of the following attributes, some questions are stated that should help you make a judgment, and a definition of an 'excellent' [\*\*\*\*] rating is provided.

Items to be rated 0 – 5 (half ratings are acceptable)

**Rating**

2.12	<b>Overall adequacy of computer generated reports:</b> (This overall rating is obtained by using judgment based on the ratings given for items 2.12.1 – 2.12.7. <b>Do not simply average numbers to obtain an overall rating.</b> )	[     ]
2.12.1	<p><b>Scope or coverage</b>            Reports can be seen as varying in both their breadth and their specificity. Reports may also vary in the range of people for whom they are suitable. In some cases it may be that separate tailored reports are provided for different groups of recipients.</p> <ul style="list-style-type: none"> <li>• <i>Does the report cover the range of attributes measured by the instrument?</i></li> <li>• <i>Does it do so at a level of specificity justifiable in terms of the level of detail obtainable from the instrument scores?</i></li> <li>• <i>Can the 'granularity' of the report (i.e. the number of distinct score bands on a scale that are used to map onto different text units used in the report) be justified in terms of the scales measurement errors?</i></li> <li>• <i>Is the report used with the same populations of people for who the instrument was designed? (e.g. Groups for whom the norm groups are relevant, or for whom there is relevant criterion data etc).</i></li> </ul> <p>[****] Excellent fit between the scope of the instrument and the scope of the report, with the level of specificity in the report being matched to the level of detail measured by the scales. Good use made of all the scores reported from the instrument.</p>	[     ]
2.12.2	<b>Reliability</b>	

	<ul style="list-style-type: none"> <li>• <i>How consistent are reports in their interpretation of similar sets of score data?</i></li> <li>• <i>If report content is varied (e.g. by random selection from equivalent text units) is this done satisfactorily?</i></li> <li>• <i>Is the interpretation of scores and differences between scores justifiable in terms of the scale measurement errors?</i></li> </ul> <p>[*****] Excellent consistency in interpretation and appropriate warnings provided for statements, interpretation and recommendations regarding their underlying errors of measurement</p>	[ ]
2.12.3	<p><b>Relevance or validity</b></p> <p>The linkage between the instrument and the content of the report may be either explained within the report or be separately documented. Where reports are based on clinical judgement, the process by which the expert(s) produced the content and the rules relating scores to content should be documented.</p> <ul style="list-style-type: none"> <li>• <i>How strong is the relationship between the content of the report and the scores on the instrument? To what degree does the report go beyond or diverge from the information provided by the instrument scores?</i></li> <li>• <i>Does the report content relate clearly to the characteristics measured by the instrument?</i></li> <li>• <i>Does it provide reasonable inferences about criteria to which we might expect such characteristics to be related?</i></li> <li>• <i>What empirical evidence provided to show that these relationships actually exist?</i></li> </ul> <p>It is relevant to consider both the construct validity of a report (i.e. the extent to which it provides an interpretation that is in line with the definition of the underlying constructs) and criterion-validity (i.e. where statements are made that can be linked back to empirical data).</p> <p>[*****] Excellent relationship between the scales and the report content, with clear justifications provided.</p>	[ ]
2.12.4	<p><b>Fairness, or freedom from systematic bias</b></p> <ul style="list-style-type: none"> <li>• <i>Is the content of the report and the language used likely to create impressions of inappropriateness for certain groups?</i></li> <li>• <i>Does the report make clear any areas of possible bias in the results of the instrument?</i></li> <li>• <i>Are alternate language forms available? If so, have adequate steps been taken to ensure their equivalence?</i></li> </ul> <p>[*****] Excellent, clear warnings and explanations of possible bias, available in all relevant user languages</p>	[ ]
2.12.5	<p><b>Acceptability</b></p> <p>This will depend a lot on the complexity of the language used in the report, the complexity of the constructs being described and the purpose for which it is intended.</p> <ul style="list-style-type: none"> <li>• <i>Is the form and content of the report likely to be acceptable to the intended recipients?</i></li> <li>• <i>Is the report written in a language that is appropriate for the likely levels of numeracy and literacy of the intended reader?</i></li> </ul> <p>[*****] Very high acceptability, well-designed and well-suited to the intended audience</p>	[ ]
2.12.6	<p><b>Practicality</b></p> <p>Practicality issues also affect acceptability. The main practical advantage of computer-generated reports is that they save time for the person who would otherwise have to produce the report. When that person is not the end-user, the practicality arguments may be harder to make.</p> <ul style="list-style-type: none"> <li>• <i>How much time does each report save the user?</i></li> <li>• <i>How much time does each report take to read and use?</i></li> </ul> <p>[*****] Excellent in terms of efficiency and value.</p>	[ ]
2.12.7	<p><b>Length</b></p> <p>This is an aspect of Practicality and should be reflected in the rating given for this. More specifically this provides an index of the ratio of quantity of output to input. The number of scales on which the report content is based are regarded as the input, and the number of report pages (excluding title pages, copyright notices etc) are regarded as the output.</p> <p><i>To calculate this index</i>, count up the number of scales, including derived scales and composite scales (e.g. for personality measures, higher order factor scales, scales for team types, leadership styles etc may be derived from the base scales).</p> <ol style="list-style-type: none"> <li>1. Divide the total number of pages by the number of scales.</li> <li>2. Multiple this ratio by 10 and round the result to the nearest integer.</li> </ol> <p>Generally values greater than 10 are likely to indicate reports that may be over long and over-interpreted.</p> <p>E.g.: Development Report - <math>8/7 \times 10 = 11.42</math>.</p>	[ ]

2.12.8

**Reviewers' comments on Computer Generated Reports:**

The evaluation can consider additional matters such as whether the reports takes account of checks on consistency of responding, response bias measures (e.g. measures of central tendency in ratings) and other indicators of the confidence with which the person's scores can be interpreted.

Comments on the complexity of the algorithms can be included. For example, whether multiple scales are considered simultaneously, how scale profiles are dealt with etc. Such complexity should, of course, be supported in the manual by a clear rationale.

Judging computer-based reports is made difficult by the fact that many suppliers will, understandably, wish to protect their intellectual property in the algorithms and scoring rules. In practice, sufficient information should be available for review purposes from the technical manual describing the development of the reporting process and its rationale, and through the running of a sample of test cases of score configurations.

Ideally the documentation should also describe the procedures that were used to test the report generation for accuracy, consistency and relevance.

---

---

**Section 8:**  
**Final Evaluation:**

---

---

3.4

**Evaluative Report of the Test:**

This section should contain a concise, clearly argued judgement about the instrument/product. It should describe its pros and cons, and give some general recommendations about how and when it might be used - together with warnings (where necessary) about when it should not be used.

The evaluation should cover topics such as the appropriateness of the instrument for various assessment functions or areas of application; any special training needs or special skills required; whether training requirements are set at the right level; ease of use; the quality and quantity of information provided by the supplier and whether there is important information which is not supplied to users.

Include comments on any research that is known to be under way, and the supplier's plans for future developments and refinements etc.

	<b>Conclusions:</b>	
4.0	<p><b>Recommendations</b> (<i>select one</i>)  The relevant recommendation, from the list given, should be indicated. Normally this will require some comment, justification or qualification. A short statement should be added relating to the situations and ways in which the instrument might be used, and warnings about possible areas of misuse.</p> <p><b>All the characteristics listed below should have ratings of either [n/a], [2], [4], [5] if an instrument is to be 'recommended' for general use (box 5 or 6).</b></p> <p>[2.9] Norms and reference groups  [2.10.1] Construct validity  [2.10.2] Criterion-related validity  [2.11] Reliability-overall  [2.12] Computer generated reports</p> <p>If any of these ratings are [ ] or a [*] the instrument will normally be classified under Recommendation 1, 2, 3, or 4 or it will be classified under 'Other' with a suitable explanation given.</p>	<input type="checkbox"/> 1 Research only tool. Not for use in practice. <input type="checkbox"/> 2 Only suitable for use by an expert user under carefully controlled conditions or in very limited areas of application <input type="checkbox"/> 3 Suitable for supervised use in the area(s) of application defined by the distributor by any user with general competence in test use and test administration <input type="checkbox"/> 4 Requires further development. Only suitable for use in research <input type="checkbox"/> 5 Suitable for use in the area(s) of application defined by the distributor, by test users who meet the distributor's specific qualifications requirements <input type="checkbox"/> 6 Suitable for unsupervised self-assessment in the area(s) of application defined by the distributor <input type="checkbox"/> 7 Other

5.	<b>Notes References and Bibliography</b>
<p>You should check the standard sources for reviews for each instrument (e.g. 'Buros' and Test Critiques). You should add details of any references cited in your Evaluation and list references to any other reviews that you know of. Where relevant, you can add a brief annotation (no more than 50 words or so) to each one concerning the conclusions drawn in that review. Indicate any other sources of information that might be helpful to the user.</p>	
See 1.10.5	<b>Constructs Measured:</b>

## **Appendix D: Competencies of Test Administrators**

Extract from Consultation Draft of European Test User Standards for test use in Work and Organizational settings



"European Test User  
Standards for test us

Extract from Consultation Draft of European Test User Standards for test use in Work and Organizational settings

<b>Unit 3</b>	<b>Follow good practice in the administration of tests</b>
<b>Standard: 3.1</b>	<b>Make necessary preparations for the testing session</b>
<b>Standard: 3.2</b>	<b>Administer the tests properly</b>
<b>Standard: 3.3</b>	<b>Score the test results accurately</b>

What is this unit about?	This unit is about following good practice in administering psychological tests
Who is this unit for?	This unit should be appropriate for all those involved in the administration of tests to individuals or groups.

<p><b>Occupational Context:</b></p> <p>Testing of people for:</p> <ul style="list-style-type: none"> <li>• Recruitment and selection</li> <li>• Organizational or team fit</li> <li>• Identification of development needs</li> <li>• Career advice and guidance</li> <li>• Other organizational purposes</li> </ul> <p>Testing of groups for:</p> <ul style="list-style-type: none"> <li>• Team development</li> <li>• Competencies audit</li> <li>• Other organizational purposes</li> </ul> <p>Testing methods</p> <ul style="list-style-type: none"> <li>• Psychological tests of ability (maximum performance measures)</li> <li>• Psychological tests of personality, motivation etc (typical performance measures)</li> </ul> <p>Administration modes</p> <ul style="list-style-type: none"> <li>• Individual vs Group</li> <li>• Paper-based, Mechanical, Computer-based</li> <li>• Online (internet) vs offline testing</li> <li>• Open, Controlled, Supervised or Managed administration</li> </ul>
--

<b>Knowledge required in this Unit</b>		
<b>Theories, models and principles of:</b>	<b>Facts, data and information about:</b>	<b>Methods, techniques and procedures relating to:</b>
<p>Ethics and principles of good practice in test administration</p> <ul style="list-style-type: none"> <li>• Informed consent</li> <li>• Privacy and confidentiality</li> </ul>	<p>Psychological tests</p> <p>Good practice guides relating to test administration:</p> <ul style="list-style-type: none"> <li>• ITC Guidelines on Test Use</li> <li>• ITC Guidelines on Computer-based testing and testing on the Internet</li> <li>• Local and national good practice guides relating to testing and test use</li> </ul> <p>Law and Codes of Practice relating to protection of personal data and intellectual property rights in relation to testing</p> <p>Law and Codes of Practice relating to equal opportunities, direct and indirect discrimination, and employment law.</p>	<p>Assessment needs analysis</p> <p>Testing of people for:</p> <ul style="list-style-type: none"> <li>• Recruitment and selection</li> <li>• Organizational or team fit</li> <li>• Identification of development needs</li> <li>• Career advice and guidance</li> <li>• Other organizational purposes</li> </ul> <p>Testing of groups for:</p> <ul style="list-style-type: none"> <li>• Team development</li> <li>• Competencies audit</li> <li>• Other organizational purposes</li> </ul>

<b>Standard: 3.1</b>		<b>Make necessary preparations for the testing session</b>
	<b>Performance required</b>	<b>Occupational Context</b>
	You must ensure that you:	Testing of people for:
<b>a</b>	Provide relevant parties in a timely manner with clear information concerning the purpose of testing, how the results will be used and data management arrangements.	<ul style="list-style-type: none"> <li>• Recruitment and selection</li> <li>• Organizational or team fit</li> <li>• Identification of development needs</li> <li>• Career advice and guidance</li> <li>• Other organizational purposes</li> </ul>
<b>b</b>	Provide guidance on ways in which test takers might best prepare for the test session, and the procedures to be followed.	Testing of groups for:
<b>c</b>	Check that the test is appropriate for the linguistic or dialectic group of the test taker.	<ul style="list-style-type: none"> <li>• Team development</li> <li>• Competencies audit</li> <li>• Other organizational purposes</li> </ul>
<b>d</b>	Send test takers approved practice, sample, or preparation materials where these are available and where this is consistent with recommended practice for the tests concerned.	Testing methods
<b>e</b>	Explain clearly to test takers their rights and responsibilities.	<ul style="list-style-type: none"> <li>• Psychological tests of ability (maximum performance measures)</li> <li>• Psychological tests of personality, motivation etc (typical performance measures)</li> </ul>
<b>f</b>	Gain the explicit informed consent of test takers or their representatives before any testing is done.	Administration modes
<b>g</b>	Explain, when testing is optional, the consequences of taking or not taking the test to relevant parties so that they can make an informed choice.	<ul style="list-style-type: none"> <li>• Individual vs Group</li> <li>• Paper-based, Mechanical, Computer-based</li> <li>• Online (internet) vs offline testing</li> <li>• Open, Controlled, Supervised or Managed administration</li> </ul>
<b>h</b>	Make the necessary practical arrangements for the test sessions.	
<b>i</b>	Ensure test centres or locations are available and suitably equipped.	
<b>j</b>	Ensure that staff who will be involved in the administration are competent.	
<b>k</b>	Inform the test taker of the testing context and conditions	
<b>l</b>	Prepare the testing methods, equipment and materials	
<b>m</b>	Make arrangements for the testing of people with disabilities	

<b>The Learning Specification</b>	
<b>Essential Knowledge</b>	<b>Essential Skills</b>
Relevant testing methods and instruments	Management of testing materials General administration

<b>Standard: 3.2 Administer the tests properly</b>	
<b>Performance required</b>	<b>Occupational Context</b>
<p>You must ensure that you:</p> <p><b>a</b> Welcome test-takers and brief them in a positive fashion, and act to reduce their anxiety.</p> <p><b>b</b> Provide appropriate assistance to test takers who show signs of undue distress or anxiety.</p> <p><b>c</b> Carry out administration procedures as specified in test manuals.</p> <p><b>d</b> Deal appropriately with any questions, technical or personal problems or issues arising during the testing session.</p> <p><b>e</b> Observe and record deviations from test procedures.</p> <p><b>f</b> Ensure the security and safety of testing materials and that all materials are accounted for at the end of each testing session</p> <p><b>g</b> Adhere strictly to the directions and instructions specified in test manuals while making reasonable accommodations for persons with disabilities.</p>	<p>Testing of people for:</p> <ul style="list-style-type: none"> <li>• Recruitment and selection</li> <li>• Organizational or team fit</li> <li>• Identification of development needs</li> <li>• Career advice and guidance</li> <li>• Other organizational purposes</li> </ul> <p>Testing of groups for:</p> <ul style="list-style-type: none"> <li>• Team development</li> <li>• Competencies audit</li> <li>• Other organizational purposes</li> </ul> <p>Testing methods</p> <ul style="list-style-type: none"> <li>• Psychological tests of ability (maximum performance measures)</li> <li>• Psychological tests of personality, motivation etc (typical performance measures)</li> </ul> <p>Administration modes</p> <ul style="list-style-type: none"> <li>• Individual vs Group</li> <li>• Paper-based, Mechanical, Computer-based</li> <li>• Online (internet) vs offline testing</li> <li>• Open, Controlled, Supervised or Managed administration</li> </ul> <p>Technical problems and issues</p> <ul style="list-style-type: none"> <li>• Computer system crashes</li> <li>• Loss of Internet connection</li> </ul> <p>Personal problems and issues</p> <ul style="list-style-type: none"> <li>• Illness or sickness</li> <li>• Anxiety</li> <li>• Session interruptions or disruptions</li> </ul>

<b>The Learning Specification</b>	
<b>Essential Knowledge</b>	<b>Essential Skills</b>
<p>The possible impact of assessment on test takers</p> <p>The possible impact of mode of administration on the quality of test results</p> <p>The possible impact of anxiety on test performance.</p>	<p>General skills associated with managing issues and problems that might arise in testing sessions</p> <p>Test administration process skills relating to different types of tests and conditions of administration (group, individual; interactive non-interactive)</p> <p>Management of testing materials</p>

<b>Standard: 3.3</b>		<b>Score test results accurately</b>
	<b>Performance required</b>	<b>Occupational Context</b>
	You must ensure that you:	Testing of people for: <ul style="list-style-type: none"> <li>• Recruitment and selection</li> <li>• Organizational or team fit</li> <li>• Identification of development needs</li> <li>• Career advice and guidance</li> <li>• Other organizational purposes</li> </ul> Testing of groups for: <ul style="list-style-type: none"> <li>• Team development</li> <li>• Competencies audit</li> <li>• Other organizational purposes</li> </ul> Testing methods <ul style="list-style-type: none"> <li>• Psychological tests of ability (maximum performance measures)</li> <li>• Psychological tests of personality, motivation etc (typical performance measures)</li> </ul> Administration modes <ul style="list-style-type: none"> <li>• Individual vs Group</li> <li>• Paper-based, Mechanical, Computer-based</li> <li>• Online (internet) vs offline testing</li> <li>• Open, Controlled, Supervised or Managed administration</li> </ul>
<b>a</b>	Score test results according to standardised procedures where scoring is not automated.	
<b>b</b>	Generate normed and derived scores according to documented procedures associated with the testing method.	
<b>c</b>	Uses procedures to screen test results to recognise improbable or unreasonable scores. Checks score scale-conversions and other clerical procedures for accuracy.	
<b>d</b>	Clearly and accurately labels scales and provides clear identification of norms, scales types, and equations used.	

<b>The Learning Specification</b>	
<b>Essential Knowledge</b>	<b>Essential Skills</b>
Test scoring procedures	Use of scoring keys and self-scoring test forms
Norms and standardisation	Use of norm tables
Aberrant response patterns	Conversion of test raw scores to standard scores
	Bringing together and documenting of tests scores and other assessment data for reports.
	Computation, where appropriate, of composite scores using standard formulae and equations.
	Management and filing of data
	Procedures for checking for clerical errors.

## Appendix E: BPS Test Review Process



"Review process  
2006v2.doc"

## ***BPS Test Review process 2006: Using the EFPA Criteria.***

The Editorial Board consists of a Senior Editor and a number of Consulting Editors. The Senior Editor and Consulting editor appointments are approved by the BPS Steering Committee on Test Standards. The appointments are part time. The Senior Editor sits ex officio on the Executive Committee of the BPS Psychological Testing Centre (PTC).

The reviewers were selected from a panel of Chartered Psychologists identified by the Editorial Board. The people selected for the panel all had expertise in psychometrics or extensive practical experience in using Level A/B tests, or both. All reviewers receive training in the use of the review criteria and process.

The review process was designed to provide informed and balanced evaluation.

- The Editorial Board select two reviewers for each instrument from the panel. Their selection is based on the following criteria: wherever possible, one will have particular psychometric expertise and the other practical experience in use of the instrument - or similar types of instruments; neither may have any direct involvement with or vested interest in the instrument - either as supplier or developer, or, indeed, as a direct competitor.
- The reviewers are provided with a standard set of review criteria and set of guidelines (i.e. The EFPA Review Criteria). They are not informed as to the identity of the other reviewer at this stage, nor are they allowed to contact the supplier of the instrument directly - all contact is through the Consulting Editor or the office.
- All those involved in producing the reviews of an instrument (i.e. the two reviewers, the Consulting Editor and Senior Editor) are required to sign the following affirmation:  
*I hereby acknowledge that I have had no involvement in the design, development and production of the instrument reviewed here, and do not stand to benefit directly from sales of it.*
- While those with a direct commercial involvement in an instrument are excluded from the process of reviewing that instrument, involvement as a reviewer of other instrument (as a practitioner or researcher) is considered in a positive light. The assignment of instruments to Consulting Editors involves similar considerations: wherever possible, each instrument is assigned to an Editor with particular knowledge and experience of it - either as practitioner, researcher or both.
- The material on which the reviews are based includes *whatever the supplier provided to the BPS as 'evidence'*. Generally, this is the instrument itself with its user and technical documentation; information about specialist training courses; key scientific papers, technical reports etc. While reviewers are free to consult other references and other reviews, the need to base the review on an evaluation of *the materials generally available to occupational users* is stressed.
- When the two reviews had been completed they are used by the Consulting Editor as the basis for the first draft of the final review. Once drafts had been completed, they are checked for consistency and edited by the Senior Editor.

- The final draft reviews are then sent to the two reviewers and the supplier for their factual comments. Thus, the supplier is provided with an opportunity to comment on matters of interpretation and to correct errors of fact in the review.
- Amendments and corrections to the draft reviews are made by the Senior Editor, with input from the relevant Consulting Editor where needed.
- The final version of the review is prepared for publication by the Senior Editor, having considered any comments provided by the reviewers, the Consulting Editors and supplier, and submitted for approval by the Editorial Board.

For the most part, there are only very minor disagreements between reviewers. These can be easily be reconciled and accommodated within the final review by the Consulting Editor.

Throughout, the Editorial Board are concerned to provide a fair balance between discussion of the weaknesses and virtues of each instrument. In particular, they strive to ensure that each instrument is evaluated with due regard to its own frame of reference and intended mode of use - notwithstanding the requirement for it to produce reliable and valid measures.

The final review is published with the names of the reviewers, consulting editor and senior editor attached. PDF reports are generated from the database by the PTC's software. These are made available to subscribers or on a pay-per-use basis.

The mechanics of this process have been greatly streamlined in the past few years by the development of a web-enabled database. Reviewers and editors work directly in this database through a content management system and the final approved version of the review is published onto the PTC website ([www.psychtesting.org.uk](http://www.psychtesting.org.uk)) from the database by the BPS office staff.

The Senior Editor is Dr Patricia Lindley.

For further details contact the PTC Manager, Julie Mrowicki: [julmro@bps.org.uk](mailto:julmro@bps.org.uk)