

## Recommendations for regulating development, control and use of psychological tests

### People Assessment in Industry (PAI)

Compiled on behalf of the Society for Industrial and Organisational Psychology of South Africa (SIOPSA), by People Assessment in Industry (PAI), an interest group of SIOPSA.



## Contents

<b>1. Introduction and background</b> .....	<b>1</b>
1.1 About People Assessment in Industry (PAI) .....	1
<b>2. General comments concerning the regulation of psychological tests</b> .....	<b>2</b>
<b>3. Specific member comments on the regulation of psychological tests</b> .....	<b>3</b>
3.1 Distinguishing between a classification and a quality evaluation process .....	3
3.2 Differentiating between 'clinical' and 'industrial' tests.....	4
3.3 Designing a modern framework that is 'future-proofed' .....	5
3.4 Sufficient training and allowing for professional judgement.....	5
3.5 Creating a stream-lined and user-friendly process for classifying tests .....	5
3.6 Competence to review tests and competitive / vested interest .....	5
<b>4. SIOPSA's recommendation for the way forward</b> .....	<b>6</b>
<b>5. In closing</b> .....	<b>7</b>
<b>6. APPENDIX A:</b> .....	<b>8</b>
<b>7. APPENDIX B:</b> .....	<b>9</b>
<b>8. APPENDIX C:</b> .....	<b>26</b>
<b>9. APPENDIX D:</b> .....	<b>34</b>

## 1. Introduction and background

The Society for Industrial and Organisational Psychology of South Africa (SIOPSA) wishes to thank the Professional Board for Psychology for the opportunity to provide inputs to the intended regulations for the development, control and use of psychological tests in South Africa. We also wish to express our commitment to working constructively with the Board on areas of mutual concern.

It is SIOPSA's aim to provide constructive comments and inputs to the proposed regulations and would like to assure the Board of a commitment to ensuring a practical and clearly defined outcome. This document has been prepared by People Assessment in Industry (PAI), an interest group of the Society for Industrial and Organisational Psychology of South Africa (SIOPSA), on behalf of our members.

The opinions expressed in this document are based on inputs from SIOPSA members across various touch-points, including:

- comments received from SIOPSA members who responded to a request for inputs circulated on the 9<sup>th</sup> December 2011 in SIOPSA's quarterly newsletter, with a follow-up electronic reminder sent out on the 19<sup>th</sup> January 2012
- opinions received from SIOPSA members who responded to an electronic communiqué circulated on 4 February 2011 requesting comments on the proposed amendments to Section 8 of the Employment Equity Act (Government Gazette No. 33873), relating to psychological and other similar assessments (refer to Appendix B).
- delegates who participated in a practitioner's forum on 9 June 2006 at the SIOPSA conference held at the CSIR in Pretoria, with the purpose of the forum being to formulate SIOPSA's comments and views to the Professional Board for Psychology's 2006 proposal concerning the reclassification of psychological tests (refer to Appendix C). Delegates at this forum comprised of psychologists registered in the Industrial, Clinical, Counselling and Research categories, working as consultants/private practitioners, as academics, and as professionals in the corporate environment.

### 1.1 *About People Assessment in Industry (PAI)*

People Assessment in Industry (PAI) was established in 1998 to proactively engage various stakeholders interested in psychological assessment in South African industry.

Functioning as an interest group of the Society for Industrial and Organisational Psychology in South Africa (SIOPSA), the purpose of PAI is to ensure testing exists in the South African context as a value adding and ethical practice. The group's activities include promoting fair assessments in the workplace, with the primary aim being to be generally recognised as the leading organisation in the continuing promotion of fair assessment in the workplace to which people will automatically refer for assessment expertise and knowledge.

Prior to the promulgation of the 1998 Employment Equity Act, PAI gave input to government on the wording as relating to psychological assessment. A specific contribution made at that stage concerning section 8 was the inclusion of the wording "Psychological testing and other similar assessments", which was done with the intention of raising the standards of assessment practices—for both psychological and non-psychological assessments—in South Africa in general.

PAI has also brought out a number of publications dedicated to providing guidance to practitioners in terms of ethical test use, including:

- Guidelines for the Validation and Use of Assessment Procedures for the Workplace (2005)
- Code of Practice for Psychological and Other Similar Assessment in the Workplace (2006)
- Issues Facing Organisations Using Assessments in the Workplace (2007).

## 2. General comments concerning the regulation of psychological tests

Individual members and SIOPSA as a professional body both acknowledge the need for control to be exercised over certain types of assessments. We believe this control can be achieved through the establishment of a considered and comprehensive test classification framework that differentiates appropriately between testing that is of a clinical nature based on a medical model, and testing in industry.

However, in recent years a number of concerns have been identified with the current state of affairs in the regulation of psychological tests in South Africa. Work by a SIOPSA Task Group at the 2006 SIOPSA conference and thereafter provides a useful overview of these concerns:

- Recognition that the Health Professions are not the only ones that use assessment tools (tests) and procedures is required. The entire field of marketing research, opinion polling, social anthropology, political science, organisational diagnosis by means of culture surveys, and many more, use instruments and assessment procedures. Some of these procedures could be classified as psychological acts. The emphasis should therefore be on the entire process of assessment, rather than on the classification of tests to facilitate the determination of the category of tester who may use the instrument.
- The purpose of the Board is to protect the interests of the public. Classification and re-classification of tests should be based on a model that recognises the complexity of psychological assessment as well as the variety in the circumstances (context) under which we utilise psychological tests. Test classification should take into account the difference between testing that is of a clinical nature based on a medical model, and testing in industry.
- The current classification procedure leaves no room for the professional judgment of the psychologist.
- The current classification of tests as being psychological or non-psychological is narrow. A clear need was expressed that further differentiation within the category of psychological tests is required. In this regard reference was made to the previous, discontinued level A, B and C test categories. The different levels should then indicate different control guidelines e.g. administration, interpretation etc.
- An integrated system is needed that includes all modes of test administration. A further directive of members was that the classification system be systemic, simple and dynamic so that South Africa is not left behind following complex procedures while the rest of the world advances in actual test development.

- In the current classification process the description and evaluation of the test is not in the public domain. Ratings should be built into the classification system to inform stakeholders of the quality of test(s). The test descriptions and evaluations done by the Psychometrics Committee should also serve a broader educational purpose by making them available in the public domain and open to debate.

### 3. Specific member comments on the regulation of psychological tests

Previous discussions on the matter of test regulation in South Africa have referred to the work of the European Federation of Psychologists' Associations (EFPA), and specifically their *Review Model for the Description & Evaluation of Psychological Tests* (see Appendix D). SIOPSA believes the model is a better differentiated format for test description and evaluation, and addresses the concerns previously raised by SIOPSA members.

In addition to the above, specific comments received from members during this round of consultation are detailed in the following sections.

#### 3.1 Distinguishing between a classification and a quality evaluation process

A suggestion received was to distinguish between the classification process – i.e. what level or type of test a particular instrument considered to be – and the quality evaluation process – i.e. does the instrument's psychometric properties fulfil requirements. This is to avoid potential confusion arising from test users assuming whether or not a particular instrument has been classified is an indication of the quality of the instrument rather than a consideration of the type of instrument and necessary training to use it.

By way of example, the EFPA model describes three levels of competence needed to practice safely in various test roles (Bartram, 2011)<sup>1</sup>:

Level	Description
Level 1	Level 1 relates to use under supervision: Competent to practice under supervision. It includes Test Administration or other qualifications relating to some part of test use. An outline has been produced for a Level 1 qualification that would cover use by line managers of reports generated by tests, for example.
Level 2	Level 2 is the qualified test user: Competent to practice independently in a limited range of situations. The Human Resources professional or Work/Organizational psychologist who uses tests in a routine way assessment for selection, development, etc., would be Level 2.
Level 3	Level 3 involves a degree of expertise in depth and/or greater breadth: Competent to practice independently in a broad range of situations with expertise in one or more areas. Psychologists who have specialized in testing and other professional experts in testing (e.g., psychometricians) could be Level 3.

The EFPA model also includes an evaluation of the norms, reliability and validity of an instrument, considered to be one of the most important sections in the review and focusing on the key psychometric quality of the test. Ratings are provided of overall adequacy as well as more detailed ratings in each section (e.g. appropriateness of norm groups for local use, sample sizes, etc.) and reviewers' comments on the ratings given are also provided.

<sup>1</sup> Bartram, D. (2011). Contributions of the EFPA Standing Committee on Tests and Testing to Standards and Good Practice. *European Psychologist*, 16(2), 149-159.

EFPA standard rating	UK review representation	Explanation
[n/a]	[n/a]	This attribute is not applicable to this instrument
[-]	[None]	Not possible to rate as no or insufficient information provided
[-1]	[*]	Inadequate
[0]	[***]	Adequate or reasonable
[1]	[****]	Good
[2]	[*****]	Excellent

(Bartram, 2011)<sup>2</sup>

Utilising a similar type of quality rating system in South Africa would provide all test users with a reference guide to the quality of an instrument in an easy-to-use format.

### 3.2 Differentiating between 'clinical' and 'industrial' tests

SIOPSA previously provided recommendations that the classification process should take into account differing applications of tests, for example differentiating between clinical/diagnostic tests and instruments used in industry. This is in line with international trends, that more test categories are needed than South Africa's current 'psychological test' and 'non-psychological test'.

Comments received from members on this point clearly highlighted that different categories of tests are required, although defining the difference between 'clinical' (broadly speaking, instruments used to help diagnose and explore pathological behaviour) and 'industrial' tests (broadly speaking, instruments used to help explore and predict performance and behaviour in a work or vocational context) was considered more challenging since there are tests that have application in both clinical and industrial settings, depending on the objectives the test user would like to achieve with the testing.

A key consideration in highlighting the distinction between clinical and industrial tests (and the ones that lie somewhere in between) is that a blanket approach to treating clinical and industrial instruments as the exactly same basket of tools (and also as the 'only' categories of tests) is either limiting (if we treat all 'industrial' tests as 'clinical') or dangerous and unsuitable (treating all projective tests and more 'clinical' instruments as 'industrial' tests).

By broadening the view from classifying the test only to include considerations of application and use, as per *EFPA's European Test User Standards for Test Use in Work & Organizational Settings* as now developed into a three-level qualification system (Bartram, 2011), whereby the requirements are defined in relation to the context, instrument, and use to which it will be put, should go a long way to address this matter.

<sup>2</sup> Bartram, D. (2011). Contributions of the EFPA Standing Committee on Tests and Testing to Standards and Good Practice. *European Psychologist*, 16(2), 149-159.

### **3.3 Designing a modern framework that is 'future-proofed'**

An important requirement raised by members concerning test classification in South Africa was that we need a framework that is modern (e.g. embracing technology in testing) without compromising the validity of the tests, which at the same time is designed in such a way that enables keeping South African practice aligned with developments in the field of testing. An important point to note is that, due to the number and complexity of variables involved in assessment, it is virtually impossible to present the profession with specific, prescriptive guidelines on practice. Feeding into the next point, suggestions previously put forward include empowering the psychologist to ensure that practice conforms to fair and ethical test use and the professional should be ultimately accountable for their use of psychological tests.

Examples of international practices concerning technology and testing which can be referred to and integrated into the new classification framework include the International Test Commission's *Test-Taker's Guide to Technology-Based Testing* (2010) and *Guidelines for Computer-based and Internet-delivered Testing* (2005).

### **3.4 Sufficient training and allowing for professional judgement**

A theme that could be identified from a number of inputs, both from this round of consultation and previously, is a preference for a broad framework that advances and broadens the use of psychological assessment rather than one that restricts practice, and also a framework that empowers psychologists and test users to use their professional judgment to make decisions about which tools to use for a particular application.

One comment in particular was that the emphasis within a regulatory framework should be on test users being sufficiently well trained and competent in psychometrics to judge the applicability of the test in a particular context (taking into account ethical considerations and labour regulations). The emphasis on training is indicative of European trends as well, whereby Muniz *et. al.* (2001)<sup>3</sup> found that the restrictions imposed on testing vary considerably from country to country, but what appeared evident was that restrictions alone are no guarantee of good practice but that some form of training requirement is also necessary.

### **3.5 Creating a stream-lined and user-friendly process for classifying tests**

In comments received it was emphasised that a more stream-lined and user-friendly process was needed for creating, validating and registering new tests in South Africa, with the Professional Board for Psychology serving as an enabling body. Suggestions given for achieving this desired efficiency include the provision of templates and guidelines for test development, as well as advice and quicker turn-around times from the Board.

### **3.6 Competence to review tests and competitive / vested interest**

Several members raised concern regarding the potential for classification and especially the quality evaluation review ratings being influenced by competitive vested commercial interests. It would be useful to be more specific about who should be conducting the ratings (for example,

---

<sup>3</sup> Muniz, J. Bartram, D., Evers, A., Boben, D., Matesic, K., Glabeke, K., Fernandez-Hermida, J.R. & Zaal, J.N. (2001). Testing Practices in European Countries. *European Journal of Psychological Assessment*, 17(3), 201-211.

an independent professional) and also include criteria required in order to qualify for being a reviewer.

As raised in a previous submission, SIOPSA is cognisant of the need for economic resources and expertise to address the task of classification and therefore recommends that an independent committee be appointed. Furthermore, a cost recovery system must only be introduced once the process of classification has been agreed upon by all stakeholders involved. Resources used by the Independent Committee as well as those utilised by external reviewers ought to determine the costing structure.

#### 4. SIOPSA's recommendation for the way forward

In terms of a way forward for aligning with international best practices and guidelines while taking into account the South African context:

- SIOPSA does not support the current classification process and recommends that the EFPA review model for the description and evaluation of psychological tests or an adaptation of it be considered as an alternative to the current classification system. This process needs to be done in consultation with all stakeholders.
- The current classification of psychological tests is too narrow and SIOPSA strongly recommends that further differentiation within the category of psychological tests be considered.
- SIOPSA further recommends that the classification process should take into account differing applications of tests, for example differentiating between clinical/diagnostic tests and instruments used in industry.
- SIOPSA recommends that ratings should be built into the classification system to inform stakeholders of the quality of tests (e.g. in terms of psychometric properties such as norm groups, validity and reliability). Ratings should be applicable when developing a new measure or adapting an existing measure for use in South Africa. Ratings on the evaluation of an instrument, based on the local standardisation and psychometric properties of the instrument, will assist psychologists in making an informed decision on the applicability of the instrument for use in South Africa.
- SIOPSA is of the opinion that a classification framework should leave some room for professional judgment.
- Define the criteria to qualify for serving as a test reviewer, to simultaneously ensure quality as well as objectivity in the classification and review process.
- Acknowledging differing levels of tests (as per the EFPA review model) and utilising this to inform test user qualification requirements can be done drawing on EFPA's (2005) *European Test User Standards for Test Use in Work & Organizational Settings*, as now developed into a three-level qualification system whereby the requirements are defined in relation to the context, instrument, and use to which it will be put.
- Finally, SIOPSA recommends that comments previously forwarded by stakeholders concerning guidelines on computerised assessments as well as existing international best practice guidelines (for example, the International Test Commission's 2010 *Test-Taker's Guide to Technology-Based Testing*) be integrated into the new classification framework to ensure South African practice remains aligned with developments in the field of testing.

## 5. In closing

It was SIOPSA's aim to provide constructive comments given the Board's communication that they are currently compiling regulations concerning the development, control and use of psychological tests, instruments and devices. SIOPSA would like to assure the Board of a commitment to ensuring a practical and clearly defined outcome, and in this regard SIOPSA would like to extend its assistance in any way possible.

In general SIOPSA would like to suggest that there be more communication on the usefulness of specific tests and classification of tests with the general public, as well as practitioners and psychologists.

We look forward to the Board's response.

Sincerely



Kim Dowdeswell

**Chair: PAI, an Interest Group of SIOPSA**

## 6. APPENDIX A:

This document was compiled based on the commentary provided by SIOPSA members in previous rounds of consultation (refer to Appendices B and C) as well inputs into the current round provided by members listed below:

Name	HPCSA Registration	Position	Organisation
Kevin Distiller	PS 0086916	Director	Latitude 26
Kim Dowdeswell	PS 0096008	Head of Products & Research	SHL South Africa
Arthur Hammond		Regional Manager	Axiom Software SA
Marcel Harper	PS 0075442	Chief Operator Officer	Omnikor
Mollie Lyons	PS 0026166	Learning & Development Manager	WorleyParsons
Hugo Misselhorn	PS 0005496	Owner	MOD Consulting
Cynthia Rose	PS 0037893	Clinical Psychologist	Private Practice
Hilton Rudnick	PS 0057533	Managing Director	Omnikor
Georgia Stylianou		Organisational Psychologist	Private Practice
Nicola Taylor		Head of Research	Jopie van Rooyen & Partners
Ralph Wortley	PS 003174	Private Practice	SIOPSA Advisory Committee

## **7. APPENDIX B:**

PAI's COMMENTARY ON THE PROPOSED AMENDMENTS TO THE EMPLOYMENT EQUITY ACT  
(FEBRUARY 2011)

## Commentary on the proposed amendments to the Employment Equity Act

### People Assessment in Industry (PAI)

Commentary on the proposed amendments to the Employment Equity Act, published for public comment on the 17th December 2010 in Government Gazette No. 33873. Compiled, on behalf of the Society for Industrial and Organisational Psychology of South Africa (SIOPSA), by People Assessment in Industry (PAI), an interest group of SIOPSA.



## Contents

<b>1. Introduction and Background.....</b>	<b>1</b>
1.1 About People Assessment in Industry (PAI) .....	1
<b>2. General Comments Regarding the Proposed Amendments .....</b>	<b>1</b>
<b>3. Specific Areas of Concern Raised Regarding the Proposed Amendments.....</b>	<b>2</b>
3.1 Section 8 of the current Act has a broader definition of assessment than the current test classification framework of the HPCSA.....	2
3.2 ‘Certification’ as per the proposed amendments as opposed to ‘Classification’ as per the HPCSA’s current test classification system.....	3
3.3 The classification system used by the HPCSA needs to be revised to account for the complexities of the modern testing environment.....	3
3.4 The application of a test must be taken into account within the greater test classification framework .....	4
3.5 The capacity and impartiality of the HPCSA .....	5
3.6 The proposed changes will not ensure that the anti-discrimination clause of the Employment Equity Act will be satisfied .....	5
<b>4. Conclusion.....</b>	<b>7</b>
<b>5. Proposal for the Way Forward.....</b>	<b>7</b>
<b>6. APPENDIX 1: .....</b>	<b>8</b>
<b>7. APPENDIX 2: .....</b>	<b>13</b>

## 1. Introduction and Background

On 17 December 2010 proposed amendments to the Employment Equity Act were published in the Government Gazette (No. 33873) for public comment. In response, this document has been compiled by People Assessment in Industry (PAI), an interest group of the Society for Industrial and Organisational Psychology of South Africa (SIOPSA), on behalf of our members.

The comments expressed in this document are based on inputs received from SIOPSA members, who responded to an electronic communiqué circulated on 4 February 2011 requesting comments on the proposed amendments. This document deals specifically with Clause 3—pertaining to Section 8 of the Act and relating to psychological and other similar assessments.

### 1.1 *About People Assessment in Industry (PAI)*

People Assessment in Industry (PAI) was established in 1998 to proactively engage various stakeholders interested in psychological assessment in South African industry.

Functioning as an interest group of the Society for Industrial and Organisational Psychology in South Africa (SIOPSA), the purpose of PAI is to ensure testing exists in the South African context as a value adding and ethical practice. The group's activities include promoting fair assessments in the workplace, with the primary aim being to be generally recognised as the leading organisation in the continuing promotion of fair assessment in the workplace to which people will automatically refer for assessment expertise and knowledge.

Prior to the promulgation of the 1998 Employment Equity Act, PAI gave input to government on the wording as relating to psychological assessment. A specific contribution made at that stage concerning section 8 was the inclusion of the wording “Psychological testing and other similar assessments”, which was done with the intention of raising the standards of assessment practices—for both psychological and non-psychological assessments—in South Africa in general.

PAI has also brought out a number of publications dedicated to providing guidance to practitioners in terms of ethical test use, including:

- Guidelines for the Validation and Use of Assessment Procedures for the Workplace (2005)
- Code of Practice for Psychological and Other Similar Assessment in the Workplace (2006)
- Issues Facing Organisations Using Assessments in the Workplace (2007).

## 2. General Comments Regarding the Proposed Amendments

The general consensus received from members was that the proposed amendments to the Employment Equity Act, specifically pertaining to the proposed amendments to Section 8 of the current Act, should be opposed, for various reasons as are laid out in the subsequent sections of this document.

At the same time, individual members and SIOPSA as a professional body acknowledge the need for control to be exercised over certain types of assessments. This control can be achieved through the establishment of a considered and comprehensive test classification framework that

differentiates appropriately between testing that is of a clinical nature based on a medical model, and testing in industry.

In the document available on the Department of Labour's website<sup>1</sup>, providing explanations of the proposed amendments, it is stated:

"Clause 3 seeks to amend section 8 of the Act to provide that only psychometric tests that have been certified by the Health Professions Council of South Africa may be used in tests and assessments of an employee. This will ensure that the testing used is scientifically valid, reliable and objective and cannot be used to unfairly discriminate and unfairly disadvantage a certain employee or a certain group of employees."

PAI acknowledges that the inclusion of the certification of psychometric tests as prerequisite for appointment of workers in the Employment Equity Amendment Bill 2010, has been done with good intentions, as the classification of psychometric assessment material should, in principle, ensure that the specific psychometric test used has been scrutinised by an independent community of experts and meets the scientific requirements in terms of presenting sufficient evidence regarding validity, reliability and culture-fairness. The inclusion of certification should also ensure that appropriately qualified test users utilise psychometric testing instruments and techniques. The proposed amendment however presupposes that the current classification process is acceptable to all stakeholders and meets best practice standards.

Furthermore, a classified psychometric test is not the only requirement for fair selection methodology. Fairness in testing depends on the complete test application, as well as the way in which information obtained through testing is used in the decision-making processes. In addition, it is unfounded to treat 'psychological tests' and 'other similar assessments' as one and the same, as they are in many ways quite different. This point is taken further in the more detailed sections that follow.

### 3. Specific Areas of Concern Raised Regarding the Proposed Amendments

#### ***3.1 Section 8 of the current Act has a broader definition of assessment than the current test classification framework of the HPCSA***

Section 8 of the Employment Equity Act currently refers to psychological tests **and other similar assessments**. Other similar assessments refer to instruments such as assessment centres, in-baskets, interviews, competency-based instruments, performance measures, etc. The current wording [correctly] points to the fact that employees can experience unfair disadvantage through decisions based on information derived from tools other than psychological tests. The current wording furthermore implies that sound psychometric properties is as important in the case of these other measuring devices as it is in the case of psychological tests.

Adding the proposed sub-paragraph (d) to section 8 as formulated below [(d) has been certified by the Health Professions Council of South Africa established in terms of the Health Professions Act, 1974 (Act No. 56 of 1974)] imposes a requirement that practically will be impossible to implement and execute. It will not be possible for the HPCSA to inspect the psychometric credentials of every interview, competency measure and performance appraisal that is currently being used in South Africa.

<sup>1</sup> EEA Explanation, available from [www.labour.gov.za/downloads/legislation/bills/proposed-amendment-bills/](http://www.labour.gov.za/downloads/legislation/bills/proposed-amendment-bills/), accessed 13/02/2011.

Furthermore, many of the tests and techniques that could fall within the blanket term ‘other similar assessments’ are not psychological in nature, which is then outside of the scope of the mandate provided by the Health Professions Act. As an example, assessment centres and simulation exercises are behavioural assessments. The current classification system only makes provision for the classification of an instrument as either psychological or not – refer to discussion under 3.2 and 3.3. Therefore, these should not be certified by the HPCSA, since they are developed on the basis of competency models. The scope of the HPCSA does not extend to the certification of competency models, so it follows that it cannot be expected of the HPCSA to determine which assessment centre can or cannot be used.

Beyond this, it must be recognised that the Health Professions (i.e. psychologists, occupational therapists and others) are not the only professions that use assessment tools (tests) and procedures. Marketing research, opinion polling, social anthropology, political science, organisational diagnosis by means of culture surveys, and many more, make use of instruments and assessment procedures.

### ***3.2 ‘Certification’ as per the proposed amendments as opposed to ‘Classification’ as per the HPCSA’s current test classification system***

It must be highlighted that the wording of the proposed amendments would bring into effect a requirement that tests be “certified by the Health Professions Council of South Africa”. Certification implies that certain standards and criteria are adhered to. Currently, however, the HPCSA only classifies tests, as per Form 208 “Policy on the Classification of Measuring Devices, Instruments, Methods and Techniques” to ensure that appropriately qualified, skilled and experienced professionals use the test, to prevent that persons are adversely affected by the test outcomes and to prevent harm to the client.

The classification of a psychometric test is done through a thorough scrutiny by experts of the capabilities and properties of the test as well as the scientific research supporting claims regarding the test. The current classification process therefore provides minimum standards to which psychological tests must adhere to, however the HPCSA has not published criteria to be used to identify and certify workplace assessment tools. In addition, the classification, description and evaluation of instruments submitted are not in the public domain. PAI and SIOPSA strongly recommend that ratings should be built into the classification system to inform stakeholders of the quality of test(s). The test descriptions and evaluations done by the Psychometrics Committee should also serve a broader educational purpose by making them available in the public domain and open to debate.

### ***3.3 The classification system used by the HPCSA needs to be revised to account for the complexities of the modern testing environment***

It is acknowledged that the use of any type of assessment—psychological or non-psychological—can have a major impact on individuals, teams and organisations. However, it has been raised before that the root of the debates in South Africa regarding the fair and ethical use of tests—including who may use tests—stems not from the nature of the user or the instrument but rather from the shortcomings of the current classification framework.

A variety of procedures, tools, techniques and guidance documents have been developed to assist Industrial and Organisational Psychologists (IOP) in effectively assessing the three primary elements in a work system – the worker, the work itself, and the work context. With regard to the assessment of worker characteristics, these procedures would include tests and other means for evaluating more stable individual differences such as cognitive abilities,

personality characteristics, values, and physical abilities. In addition, numerous procedures have been developed for analysing the content and human requirements of work, collectively referred to as job analysis procedures. With respect to the evaluation of work context variables, procedures have been developed to assess and effectively manage organisational culture and climate, organisational reward systems, and the design of organisations. These assessments will fall within the category of occupational assessments and the level of the assessments will determine the category of user.

However, the current classification system of psychological tests in South Africa is inadequate as it does not make provision for differing levels of complexity and does not cater sufficiently for all types and applications of assessment measures (e.g. simulation exercises, 360 questionnaires, assessment centres, etc.). Some psychological tests, such as projective techniques, are more complex to use while other psychological tests are highly structured and can for example be administered online without supervision. Furthermore, it must be noted that section 8 of the Employment Equity Act refers to 'Psychological **and other similar assessments**' and not only to psychological assessments, as has already been discussed earlier in this document.

SIOPSA submitted comprehensive comments to the HPCSA on 14 July 2006 (Attached hereto as Appendix 1), concerning identified shortcomings in the current classification system as well as proposed amendments and recommendations for improvements in the criteria for classification. It is strongly recommended that the current classification system be revisited by the HPCSA in collaboration with relevant stakeholders (e.g. SIOPSA, ATP SA, Universities and industry representatives) to arrive at a solution that provides a more robust approach to test classification and evaluation.

Feedback from members during the compilation of this document highlighted the importance that the test classification system should focus not only on the category of testers (as is currently the case), but more broadly in terms the purpose of the assessment, the administration and scoring of the assessment, the interpretation and feedback as well as the application of results. This is in line with previous recommendations made by SIOPSA, and other bodies, that the current test classification framework in place in South Africa should be revised to bring it more in line with international standards as used by the British Psychological Society (BPS) and the European Federation of Psychologists' Associations (EFPA). Aspects of this point—specifically that the *application* of tests as part of the definition of their use—should be taken into account in the revised test classification system is elaborated on further in Section 3.4.

### ***3.4 The application of a test must be taken into account within the greater test classification framework***

When considering the certification—or classification—of tests and assessment devices, the different aims of using instruments to gather information about people must be distinguished between. To say that only certified tests may be used is a global statement which will undermine information gathering, diagnosis and interventions by psychologists. Since information gathering and diagnosis are essential before one can offer interventions or solutions, there must be a clear distinction between instruments used for different purposes. There are instruments which provide:

1. Information and diagnosis of a person's personality or behaviour which is aimed at classifying defined personality or behaviour against established population norms for the purposes of selection for a job where criteria for successful selection have been defined.

2. Information and diagnosis of a person's personality or behaviour which is aimed at providing feedback to them for the purposes of raising self awareness aimed to behaviour, performance and attitude change.
3. Information and diagnosis of group or organization behaviour, climate/culture which is aimed at providing feedback to the group or organization to raise awareness aimed at behaviour, performance and attitude change.
4. Information and diagnosis of behaviour problems to help discern appropriate interventions for change and improvement.

Against this background a strong point was made by members that the classification framework should emphasise professional judgment as the key to legal compliance. It is however not the purpose of this document to provide detailed comments on what should be included in a suitable classification system – for a detailed discussion on proposed amendments and recommendations for improvements of the current classification system refer to Appendix 1.

### **3.5 The capacity and impartiality of the HPCSA**

Aside from the complexities regarding catering for both psychological and other similar assessments as well as the purpose of the application, several members raised concerns regarding the capacity and resources of the HPCSA to carry out the certification of all instruments and to maintain the list of certified instruments.

Furthermore, concern was raised by members regarding a conflict of interest on the HPCSA's side. The Board [Professional Board for Psychology, HPCSA] cannot act as player and referee at the same time. They have a strong business interest (classification fees) in the certification of tests. SIOPSA is cognisant of the need for economic resources and expertise to address the task of classification and therefore recommends that an independent committee be appointed. Furthermore, a cost recovery system must only be introduced once the process of classification has been agreed upon by all stakeholders involved. Resources used by the Independent Committee as well as those utilised by external reviewers ought to determine the costing structure.

### **3.6 The proposed changes will not ensure that the anti-discrimination clause of the Employment Equity Act will be satisfied**

It is further argued that even if the resources could be found to enable the HPCSA to inspect the psychometric credentials of every non-psychological assessment device used in South Africa it will not ensure that the anti-discrimination clause of the Employment Equity Act is satisfied.

The Employment Equity Act (Republic of South Africa, 1998, p.14) prohibits unfair discrimination by stating that:

No person may unfairly discriminate, directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language and birth.

It is impossible to give an assurance that the use of any assessment tool will serve the objective of the Act of "promoting equal opportunity and fair treatment in employment through the elimination of unfair discrimination" (Republic of South Africa, 1998, p. 12).

Psychological tests but also other similar assessments render information that is used in employment practices to inform decisions about employees. The manner in which this information is used to inform the decision determines whether the decision discriminates fairly or unfairly between employees. Although it is true that the use of [construct] invalid, unreliable and biased measures probably will result in decisions that unfairly discriminate against employees, the opposite is not true. Decisions based on valid, reliable unbiased information on/measures of constructs relevant to the decision can still unfairly disadvantage employees from constitutionally protected groups.

The Employment Equity Act (Republic of South Africa, 1998) is correct in describing the use of invalid predictors as an unacceptable practice since it violates the fundamental principle of the unqualified individualism position that applicants with an equal probability of succeeding on the job should have an equal probability of obtaining the job, irrespective of group membership. Since the use of a completely invalid predictor is synonymous to random selection, it gives all applicants the same probability of obtaining the job despite the fact that they differ in terms of the probability of succeeding on the job. The use of a predictor that demonstrates predictive validity, however, is not a sufficient condition to ensure that the fundamental principle comprising unqualified individualism is complied with. Even when a predictor demonstrates predictive validity, (indirect) discrimination can still unfairly disadvantage members of specific subgroups if group membership significantly explains variance in the criterion, which is not explained by the predictor, and if the selection strategy fails to take this fact into account.

Translated into psychometric terms, the Employment Equity Act (Republic of South Africa, 1998, p. 14) seems to hold the questionable position that it is not unfair discrimination to distinguish between, exclude or prefer any person on the basis of the scores obtained on a valid selection instrument. The very essence of selection is to distinguish between, exclude or show preference for individuals on the basis of measures that are systematically related to the criterion [i.e., valid selection instruments]. The question nonetheless remains whether the criterion-referenced inferences derived from the relevant predictor information does not unfairly burden or disadvantage members of specific subgroups? The definition of discrimination<sup>2</sup> provided by the Promotion of Equality and Prevention of Unfair Discrimination Act (Republic of South Africa, 2000) read in conjunction with the Cleary (Cleary, 1968) interpretation of unfair discrimination attests to the questionable nature of the Employment Equity Act position:

(viii) "discrimination" means any act or omission, including a policy, law, rule, practice, condition or situation which directly or indirectly-

- a) imposes burdens, obligations or disadvantage on; or
- b) withholds any benefits, opportunities or advantages from, any person on one or more of the prohibited grounds

If group membership does significantly explain variance in the criterion, which is not explained by the predictor, and if the selection strategy fails to take this fact into account, significant systematic group-related prediction errors will occur and the selection decision-rule will therefore discriminate unfairly since it will disadvantage members of a specific group by placing them inappropriately low in the selection rank order even though the predictor significantly correlates with the criterion.

---

<sup>2</sup> Discrimination, in terms of this definition, should not be equated with unfair discrimination but rather with adverse impact.

It is therefore argued that if amendments to the Employment Equity Act are to be considered, a far more psychometrically penetrating analysis of the current wording of the Act is required that takes paragraph 8 beyond its current argument that ensuring psychometrically sound instruments will be sufficient to ensure that nobody is unfairly discriminated against.

#### **4. Conclusion**

This document lays out the position of SIOPSA, as compiled by its interest group PAI, in terms of the proposed amendments to the Employment Equity Act, and specifically pertaining to Clause 3 (as relating to Section 8 of the current Act).

SIOPSA opposes the proposed amendments to section 8 of the Employment Equity Act as the promulgation of the proposed amendments is anticipated to result in an unsustainable situation as the current classification systems is not supported by all stakeholders and is insufficient to cater appropriately for psychological testing and “other similar assessments”. SIOPSA acknowledges the need for control to be exercised over certain types of assessments. Similar to previous engagements with the HPCSA and Professional Board for Psychology on this topic, SIOPSA believes that this control can be achieved through the establishment of a considered and comprehensive test classification framework that differentiates appropriately between the different types of assessments, their uses, and subsequent users, administrators and the like.

#### **5. Proposal for the Way Forward**

SIOPSA proposes that the proposed amendments to section 8 of the Employment Equity Act be discarded, and that the Professional Board for Psychology, on behalf of the HPCSA, engages with all relevant stakeholders to arrive at a rationale and considered test classification framework that not only meets the Board’s objective of guiding the profession while protecting the public, but that is also practical and sustainable from an implementation perspective, and that is in line with similar international standards.

## 6. APPENDIX 1:

### RE-CLASSIFICATION OF PSYCHOLOGICAL TESTS ON THE CURRENT LIST OF CLASSIFIED TESTS

14 July 2006

The Society for Industrial and Organisational Psychology (SIOPSA) wishes to thank the Professional Board for Psychology for the opportunity to comment on the proposed test re-classification process. We also wish to express our commitment to working constructively with the Board on areas of mutual concern.

SIOPSA's comments are based on inputs from delegates who participated in a practitioner's forum on 9 June 2006 at the SIOPSA conference held at the CSIR in Pretoria.

Delegates at this forum comprised of psychologists registered in the Industrial, Clinical, Counselling and Research categories, working as consultants/private practitioners, as academics, and as professionals in the corporate environment. After extensive discussions the more than hundred Industrial and Organisational Psychologists who attended the practitioner's forum unanimously expressed that they were dissatisfied with the current test classification system.

Against this background a task group was formed and mandated to:

- Investigate current international best practices in classifying psychological tests
- Open up for comment the South African guidelines on computerised testing

#### 1. CURRENT CONTEXT

Section 37 (2) (a) (b) (c) (d) and (e) of the Health Professions Act (56/1974) states that all tests that measure a psychological construct must be used, interpreted and controlled by psychologists. It became necessary to classify tests, mainly because of the need to determine boundaries around the levels of tests that different categories of testers may use. The Psychometrics Committee of the Professional Board for Psychology has been mandated to classify measuring devices, instruments, methods, techniques or tests. Subsequent to the classification of tests the Psychometrics Committee has also taken on the task of evaluating whether tests submitted for classification meets certain minimum standards as set out in Form 208. The last step in the process of classification is the issue of a certificate indicating the category of tester who may use the classified instrument.

##### 1.1 SIOPSA's position

The task group solicited opinions from SIOPSA members and identified the following concerns:

- The recognition that the Health Professions are not the only ones that use assessment tools (tests) and procedures. The entire field of marketing research, opinion polling, social anthropology, political science, organisational diagnosis by means of culture surveys, and many more, use instruments and assessment procedures. Some of these procedures could be classified as psychological acts. The emphasis should therefore be on the entire process of assessment, rather than on the classification of tests to facilitate the determination of the category of tester who may use the instrument.
- The current focus of the Psychometrics Committee is on controlling the actions of psychologists. However the purpose of the Board is to protect the interests of the public. Classification and re-classification should be based on a model that recognises the complexity of psychological assessment as well as the variety in the circumstances (context) under which we utilise psychological tests. Test classification should take into account the difference between testing that is of a clinical nature based on a medical model, and testing in industry.

- The current classification procedure leaves no room for professional judgment.
- The current classification of tests as being psychological or non-psychological is narrow. A clear need was expressed that further differentiation within the category of psychological tests is required. In this regard reference was made to the previous, discontinued level A, B and C test categories. The different levels should then indicate different control guidelines e.g. administration, interpretation etc.
- Members raised serious concerns regarding the piecemeal approach currently employed by the Professional Board in addressing strategic issues. Specific reference was made to the new practice framework, the re-classification of psychological tests, guidelines on computerised assessments as well as the policy on undesirable business practices. From a systems perspective they are all integrated and should be addressed simultaneously.
- An integrated system is needed that includes all modes of test administration. A further directive of members was that the classification system be systemic, simple and dynamic so that South Africa is not left behind following complex procedures while the rest of the world advances in actual test development.
- In the current classification process the description and evaluation of the test is not in the public domain. Ratings should be built into the classification system to inform stakeholders of the quality of test(s). The test descriptions and evaluations done by the Psychometrics Committee should also serve a broader educational purpose by making them available in the public domain and open to debate.

## **2. COMMENTS ON THE PROPOSED PROCESS OF RE-CLASSIFICATION**

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

SIOPSA recognises and supports the need for the classification of psychological tests but is not in agreement with the content and process of re-classification as proposed in the document open for discussion and consultation. SIOPSA recommends that the current classification system in its totality be reviewed and be replaced by a classification system that supports best practice and addresses the concerns raised by SIOPSA members.

Herewith detailed comments on the proposed process of re-classification:

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

SIOPSA agrees that historically classified tests which have never been reviewed and evaluated must be subjected to a classification process. However SIOPSA does not agree that the re-classification should be based on the current classification system. Historically classified instruments should be considered a priority within a new classification process as they have not been evaluated since the promulgation of the Employment Equity Act. They may not conform to its provisions.

#### **2.1.2 Need to accommodate newly created registration categories**

The need to accommodate newly created registration categories ought to be reviewed in conjunction with the new practice framework for psychology. SIOPSA is generally concerned about the Counsellor category. In particular, the demarcation between the role of Industrial Psychologist, Registered Psychological Counsellor and Psychometrist (Independent Practice) as their work roles in industry and organisations are not clear. The classification of tests can provide some clarity with regard to the role and competencies of these newly created categories as well as on the existing category of Psychological Assistant with a three year degree in psychology. Currently there is a recommendation by the Board to scrap the category of Psychological Assistant. SIOPSA, however, is of the opinion that this recommendation may be premature as the differentiation between instruments used in the clinical and work environment may necessitate retaining such a category.

Current classification certificates do not take these newly created registration categories into account. This approach may be viewed as short sighted as it does not recognise the complexities of psychological assessment as well as the variety of contexts within which psychological tests are administered. The method of classification that would be suitable must be able to define the set of competencies required in relation to the context, the instrument as well as the purpose of the assessment. In this regard SIOPSA proposes that the current system of classification be revised to accommodate a differentiation between clinical diagnostic instruments and occupational instruments used in industry.

A variety of procedures, tools, techniques and guidance documents have been developed to assist Industrial and Organisational Psychologists (IOP) in effectively assessing the three primary elements in a work system – the worker, the work itself, and the work context. With regard to the assessment of worker characteristics, these procedures would include tests and other means for evaluating more stable individual differences such as cognitive abilities, personality characteristics, values, and physical abilities. In addition, numerous procedures have been developed for analysing the content and human requirements of work, collectively referred to as job analysis procedures. With respect to the evaluation of work context variables, procedures have been developed to assess and effectively manage organisational culture and climate, organisational reward systems, and the design of organisations. These assessments will fall within the category of occupational assessments and the level of the assessments will determine the category of user.

### 2.1.3 Ensuring consistency, clarity and legal compliance

SIOPSA supports in principle that a classification system must ensure consistency, clarity and legal compliance. However we are of the opinion that the current classification system, and even a re-classification using the current format, will not be able to meet these value statements – refer to the concerns raised by SIOPSA members under point 1.1.

Due to the complexity of the assessment context a classification system should emphasise professional judgment as the key to legal compliance. Clarity and consistency can only be obtained if an all inclusive process is followed. In this regard reference can be made to the new practice framework for psychology, guidelines on computerised assessments and the policy on undesirable business practices. All these aspects are linked: from a systems perspective the one impact on the other and neither one should be addressed in isolation.

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

SIOPSA agrees that tests must be adequately supported and updated. The classification and evaluation process should however be transparent and open to public scrutiny to ensure the compliance to validity, reliability and bias as demanded by the Employment Equity Act.

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

### 2.2.1 Comment phase

SIOPSA's current submission is part of this phase. However concerns were raised by members regarding the outcome of comments received by different stakeholder groups. SIOPSA recommends that the Board integrates all comments received and communicates how the recommendations will impact the current proposal. This should be communicated to all stakeholder groups.

### 2.2.2 Review and classification

The Psychometrics Committee of the Professional Board for Psychology advanced some reasons why the re-classification of psychological tests is indicated. In all these reasons test classification is advanced as a means to support sound professional conduct and as a way to

protect the public. The proposal by the Psychometrics Committee on re-classification using the current system however presupposes that the current classification framework is acceptable to all stakeholder groups.

The task group, mandated by the members of SIOPSA, reviewed the European Federation of Psychological Association's (EFPA) model for the description and evaluation of psychological tests (Attached as Annexure 1). The task group believes the model is a better differentiated format for test description and evaluation, and addresses the concerns raised by SIOPSA members.

Dr Louise Holman, a task group member, integrated all the concerns raised by SIOPSA members and developed a South African model for the classification of psychological tests (Attached as Annexure 2). The EFPA review model has been consulted and cross referenced in the proposed model.

### 2.2.3 Order of priority

The classification of tests should be conducted in the following order of priority:

1. Historically classified tests that have never been reviewed.
2. Tests that have never been classified before.
3. Tests that have been classified under the current system – in order that they were classified.
4. Test with certificates that have been identified as anomalous or in conflict with legislation. This category will fall away if a new classification system is introduced as all current classified tests as well as new submissions will be handled as third priority

### 2.2.4 Payment

SIOPSA is cognisant of the need for economic resources and expertise to address the task of re-classifying psychological instruments. SIOPSA therefore recommends that a cost recovery system be introduced once the process of classification has been agreed upon by all stakeholders involved. Resources used by the Psychometrics Committee as well as those utilised by external reviewers ought to determine the costing structure.

## 3. SIOPSA's RECOMMENDATIONS

In summary, the above discussions indicate the following:

- SIOPSA does **not** support the current classification process and therefore recommends that the EFPA review model for the description and evaluation of psychological tests or an adaptation of it be considered as an alternative to the current classification system (Refer to Annexure 1 and 2). This process needs to be done in consultation with all stakeholders.
- The current classification of psychological tests is too narrow and SIOPSA strongly recommends that further differentiation within the category of psychological tests be considered.
- SIOPSA further recommends that the classification process should differentiate between clinical/diagnostic tests and instruments used in industry.
- SIOPSA recommends that ratings should be built into the classification system to inform stakeholders of the quality of test(s). Ratings should be applicable when developing a new measure or adapting an existing measure for use in South Africa. Ratings on the evaluation of an instrument, based on the local standardisation and psychometric properties of the instrument, will assist psychologists in making an informed decision on the applicability of the instrument for use in South Africa.

- SIOPSA is of the opinion that a classification framework should leave some room for professional judgment.
- SIOPSA recommends that the current re-classification process be put on hold pending the development of a new, revised classification process that takes into account all concerns raised by Stakeholders.
- Finally, SIOPSA recommends that comments forwarded by stakeholders on the new practice framework, guidelines on computerised assessments as well as the re-classification process be integrated within the new classification framework.

### **In closing**

It was SIOPSA's aim to provide constructive comments given the proposed process of re-classification and would like to assure the Board of a commitment to ensuring a practical and clearly defined outcome. In this regard SIOPSA would like to extend its assistance in any way possible.

In general SIOPSA would like to suggest that there be more communication on the usefulness of specific tests and classification of tests with the general public, as well as practitioners and psychologists.

We look forward to the Board's response.

Sincerely



Aletta Odendaal  
SIOPSA Task Group Co-ordinator  
SIOPSA: Past Chair

**7. APPENDIX 2:**

This document was compiled based on the inputs and commentary provided by members of SIOPSA, listed below.

<b>Name</b>	<b>HPCSA Registration</b>	<b>Position</b>	<b>Organisation</b>
Estelle Crafford	PS 002371	Manager: Employment Services	Department of Labour
Francois de Kock	PS 0085456	Lecturer: Department of Industrial Psychology	University of Stellenbosch
Kim Dowdeswell	PS 0096008	Principal Consultant & Industrial Psychologist	SHL South Africa
Hazel Dunbar	PS 0098019	Organisational Psychologist	Psychology at Work
Hilde Hay	PS 0087432	Consultant- Career & Team Development	Private Practice
Dr Louise Holman	PS 0020478	Independent Practitioner	Independent Practitioner
Illana Human	PS 0084166	Senior IPS Specialist Talent Management	Kumba Iron Ore, a member of the Anglo American plc group
Magriet Malan	PS 0054585	DD: Career Services	Department of Defence
Wisani Matsimbi	PS 0062790	Assistant General Manager: Organisational Development	South African Reserve Bank
Hugo Misselhorn	PS 0005496	Owner	MOD Consulting
Kathy Morgan	PMT 0078239	Psychometrist (Independent Practice)	Kathy Morgan Consulting
Aletta Odendaal	PS 0066389	Senior Lecturer: Industrial Psychology & People Management Independent Practitioner	University of Johannesburg Odendaal consulting
Prof Gerhard Roodt	PS 0018481	Professor: Industrial Psychology & People management	University of Johannesburg

COMMENTARY ON PROPOSED AMENDMENTS TO THE EMPLOYMENT EQUITY ACT

<b>Name</b>	<b>HPCSA Registration</b>	<b>Position</b>	<b>Organisation</b>
Prof Callie Theron		Associate Professor: Department of Industrial Psychology	University of Stellenbosch
Dr Ralph Wortley	PS 003174	Private Practice	SIOPSA Advisory Committee

## **8. APPENDIX C:**

SIOPSA'S RECOMMENDATIONS CONCERNING THE RE-CLASSIFICATION OF PSYCHOLOGICAL TESTS ON THE CURRENT LIST OF CLASSIFIED TESTS (JUNE 2006)

## **RE-CLASSIFICATION OF PSYCHOLOGICAL TESTS ON THE CURRENT LIST OF CLASSIFIED TESTS**

The Society for Industrial and Organisational Psychology (SIOPSA) wishes to thank the Professional Board for Psychology for the opportunity to comment on the proposed test re-classification process. We also wish to express our commitment to working constructively with the Board on areas of mutual concern.

SIOPSA's comments are based on inputs from delegates who participated in a practitioner's forum on 9 June 2006 at the SIOPSA conference held at the CSIR in Pretoria.

Delegates at this forum comprised of psychologists registered in the Industrial, Clinical, Counselling and Research categories, working as consultants/private practitioners, as academics, and as professionals in the corporate environment. After extensive discussions the more than hundred Industrial and Organisational Psychologists who attended the practitioner's forum unanimously expressed that they were dissatisfied with the current test classification system.

Against this background a task group was formed and mandated to:

- Investigate current international best practices in classifying psychological tests
- Open up for comment the South African guidelines on computerised testing

### **1. CURRENT CONTEXT**

Section 37 (2) (a) (b) (c) (d) and (e) of the Health Professions Act (56/1974) states that all tests that measure a psychological construct must be used, interpreted and controlled by psychologists. It became necessary to classify tests, mainly because of the need to determine boundaries around the levels of tests that different categories of testers may use. The Psychometrics Committee of the Professional Board for Psychology has been mandated to classify measuring devices, instruments, methods, techniques or tests. Subsequent to the classification of tests the Psychometrics Committee has also taken on the task of evaluating whether tests submitted for classification meets certain minimum standards as set out in Form 208. The last step in the process of classification is the issue of a certificate indicating the category of tester who may use the classified instrument.

#### **1.1 SIOPSA's position**

The task group solicited opinions from SIOPSA members and identified the following concerns:

- The recognition that the Health Professions are not the only ones that use assessment tools (tests) and procedures. The entire field of marketing research, opinion polling, social anthropology, political science,

organisational diagnosis by means of culture surveys, and many more, use instruments and assessment procedures. Some of these procedures could be classified as psychological acts. The emphasis should therefore be on the entire process of assessment, rather than on the classification of tests to facilitate the determination of the category of tester who may use the instrument.

- The current focus of the Psychometrics Committee is on controlling the actions of psychologists. However the purpose of the Board is to protect the interests of the public. Classification and re-classification should be based on a model that recognises the complexity of psychological assessment as well as the variety in the circumstances (context) under which we utilise psychological tests. Test classification should take into account the difference between testing that is of a clinical nature based on a medical model, and testing in industry.
- The current classification procedure leaves no room for professional judgment.
- The current classification of tests as being psychological or non-psychological is narrow. A clear need was expressed that further differentiation within the category of psychological tests is required. In this regard reference was made to the previous, discontinued level A, B and C test categories. The different levels should then indicate different control guidelines e.g. administration, interpretation etc.
- Members raised serious concerns regarding the piecemeal approach currently employed by the Professional Board in addressing strategic issues. Specific reference was made to the new practice framework, the re-classification of psychological tests, guidelines on computerised assessments as well as the policy on undesirable business practices. From a systems perspective they are all integrated and should be addressed simultaneously.
- An integrated system is needed that includes all modes of test administration. A further directive of members was that the classification system be systemic, simple and dynamic so that South Africa is not left behind following complex procedures while the rest of the world advances in actual test development.
- In the current classification process the description and evaluation of the test is not in the public domain. Ratings should be built into the classification system to inform stakeholders of the quality of test(s). The test descriptions and evaluations done by the Psychometrics Committee should also serve a broader educational purpose by making them available in the public domain and open to debate.

## **2. COMMENTS ON THE PROPOSED PROCESS OF RE-CLASSIFICATION**

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

SIOPSA recognises and supports the need for the classification of psychological tests but is not in agreement with the content and process of re-classification as proposed in the document open for discussion and consultation. SIOPSA recommends that the current classification system in its totality be reviewed and be replaced by a classification system that supports best practice and addresses the concerns raised by SIOPSA members.

Herewith detailed comments on the proposed process of re-classification:

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

SIOPSA agrees that historically classified tests which have never been reviewed and evaluated must be subjected to a classification process. However SIOPSA does not agree that the re-classification should be based on the current classification system. Historically classified instruments should be considered a priority within a new classification process as they have not been evaluated since the promulgation of the Employment Equity Act. They may not conform to its provisions.

#### 2.1.2 Need to accommodate newly created registration categories

The need to accommodate newly created registration categories ought to be reviewed in conjunction with the new practice framework for psychology. SIOPSA is generally concerned about the Counsellor category. In particular, the demarcation between the role of Industrial Psychologist, Registered Psychological Counsellor and Psychometrist (Independent Practice) as their work roles in industry and organisations are not clear. The classification of tests can provide some clarity with regard to the role and competencies of these newly created categories as well as on the existing category of Psychological Assistant with a three year degree in psychology. Currently there is a recommendation by the Board to scrap the category of Psychological Assistant. SIOPSA, however, is of the opinion that this recommendation may be premature as the differentiation between instruments used in the clinical and work environment may necessitate retaining such a category.

Current classification certificates do not take these newly created registration categories into account. This approach may be viewed as short sighted as it does not recognise the complexities of psychological assessment as well as the variety of contexts within which psychological tests are administered. The method of classification that would be suitable must be able to define the set of competencies required in relation to the context, the instrument as well as the purpose of the assessment. In this regard SIOPSA proposes that the current system of classification be revised to accommodate a differentiation between clinical diagnostic instruments and occupational instruments used in industry.

A variety of procedures, tools, techniques and guidance documents have been developed to assist Industrial and Organisational Psychologists (IOP) in effectively assessing the three primary elements in a work system – the worker, the work itself, and the work context. With regard to the assessment of worker characteristics, these procedures would include tests and other means for evaluating more stable individual differences such as cognitive abilities, personality characteristics, values, and physical abilities. In addition, numerous procedures have been developed for analysing the content and human requirements of work, collectively referred to as job analysis procedures. With respect to the evaluation of work context variables, procedures have been developed to assess and effectively manage organisational culture and climate, organisational reward systems, and the design of organisations. These assessments will fall within the category of occupational assessments and the level of the assessments will determine the category of user.

### 2.1.3 Ensuring consistency, clarity and legal compliance

SIOPSA supports in principle that a classification system must ensure consistency, clarity and legal compliance. However we are of the opinion that the current classification system, and even a re-classification using the current format, will not be able to meet these value statements – refer to the concerns raised by SIOPSA members under point 1.1.

Due to the complexity of the assessment context a classification system should emphasise professional judgment as the key to legal compliance. Clarity and consistency can only be obtained if an all inclusive process is followed. In this regard reference can be made to the new practice framework for psychology, guidelines on computerised assessments and the policy on undesirable business practices. All these aspects are linked: from a systems perspective the one impact on the other and neither one should be addressed in isolation.

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

SIOPSA agrees that tests must be adequately supported and updated. The classification and evaluation process should however be transparent and open to public scrutiny to ensure the compliance to validity, reliability and bias as demanded by the Employment Equity Act.

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

### 2.2.1 Comment phase

SIOPSA's current submission is part of this phase. However concerns were raised by members regarding the outcome of comments received by different stakeholder groups. SIOPSA recommends that the Board integrates all comments received

and communicates how the recommendations will impact the current proposal. This should be communicated to all stakeholder groups.

### 2.2.2 Review and classification

The Psychometrics Committee of the Professional Board for Psychology advanced some reasons why the re-classification of psychological tests is indicated. In all these reasons test classification is advanced as a means to support sound professional conduct and as a way to protect the public. The proposal by the Psychometrics Committee on re-classification using the current system however presupposes that the current classification framework is acceptable to all stakeholder groups.

The task group, mandated by the members of SIOPSA, reviewed the European Federation of Psychological Association's (EFPA) model for the description and evaluation of psychological tests (Attached as Annexure 1). The task group believes the model is a better differentiated format for test description and evaluation, and addresses the concerns raised by SIOPSA members.

Dr Louise Holman, a task group member, integrated all the concerns raised by SIOPSA members and developed a South African model for the classification of psychological tests (Attached as Annexure 2). The EFPA review model has been consulted and cross referenced in the proposed model.

### 2.2.3 Order of priority

The classification of tests should be conducted in the following order of priority:

1. Historically classified tests that have never been reviewed.
2. Tests that have never been classified before.
3. Tests that have been classified under the current system – in order that they were classified.
4. Test with certificates that have been identified as anomalous or in conflict with legislation. This category will fall away if a new classification system is introduced as all current classified tests as well as new submissions will be handled as third priority

### 2.2.4 Payment

SIOPSA is cognisant of the need for economic resources and expertise to address the task of re-classifying psychological instruments. SIOPSA therefore recommends that a cost recovery system be introduced once the process of classification has been agreed upon by all stakeholders involved. Resources used by the Psychometrics Committee as well as those utilised by external reviewers ought to determine the costing structure.

### 3. SIOPSA's RECOMMENDATIONS

In summary, the above discussions indicate the following:

- SIOPSA does **not** support the current classification process and therefore recommends that the EFPA review model for the description and evaluation of psychological tests or an adaptation of it be considered as an alternative to the current classification system (Refer to Annexure 1 and 2). This process needs to be done in consultation with all stakeholders.
- The current classification of psychological tests is too narrow and SIOPSA strongly recommends that further differentiation within the category of psychological tests be considered.
- SIOPSA further recommends that the classification process should differentiate between clinical/diagnostic tests and instruments used in industry.
- SIOPSA recommends that ratings should be built into the classification system to inform stakeholders of the quality of test(s). Ratings should be applicable when developing a new measure or adapting an existing measure for use in South Africa. Ratings on the evaluation of an instrument, based on the local standardisation and psychometric properties of the instrument, will assist psychologists in making an informed decision on the applicability of the instrument for use in South Africa.
- SIOPSA is of the opinion that a classification framework should leave some room for professional judgment.
- SIOPSA recommends that the current re-classification process be put on hold pending the development of a new, revised classification process that takes into account all concerns raised by Stakeholders.
- Finally, SIOPSA recommends that comments forwarded by stakeholders on the new practice framework, guidelines on computerised assessments as well as the re-classification process be integrated within the new classification framework.

#### **In closing**

It was SIOPSA's aim to provide constructive comments given the proposed process of re-classification and would like to assure the Board of a commitment to ensuring a practical and clearly defined outcome. In this regard SIOPSA would like to extend its assistance in any way possible.

In general SIOPSA would like to suggest that there be more communication on the usefulness of specific tests and classification of tests with the general public, as well as practitioners and psychologists.

We look forward to the Board's response.

Sincerely

A handwritten signature in black ink, appearing to read 'Aletta Odendaal'. The signature is fluid and cursive, with the first letter 'A' being particularly large and stylized.

Aletta Odendaal  
SIOPSA Task Group Co-ordinator  
SIOPSA: Past Chair

## **9. APPENDIX D:**

EFPA REVIEW MODEL FOR THE DESCRIPTION AND EVALUATION OF PSYCHOLOGICAL TESTS /  
TEST REVIEW FORM AND NOTES FOR REVIEWERS

# 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	<p><b>Description of the populations for which the test is intended:</b></p> <p>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.</p>	
1.10.5	<p><b>Number of scales and brief description of the variable or variables measured by the instrument</b></p> <p>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.</p>	
1.11	<p><b>Items format</b> (<i>select one</i>)</p> <p>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.</p>	<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	<p><b>No of Test Items:</b></p> <p>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.</p>	

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>	<p>Preparation: <input type="text"/></p> <p>Administration: <input type="text"/></p> <p>Scoring: <input type="text"/></p> <p>Analysis: <input type="text"/></p> <p>Feedback: <input type="text"/></p>

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></p> <p>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>)</p> <p>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>