Test Review

Emotional Quotient Inventory

(EQ-i 2.0)

Test Review of Emotional Quotient Inventory 2.0

Reviewers: Rachel Frost & Anna Naumenko

Consultant Editor: Roy Childs

Senior Editor: Ian Newcombe

GENERAL INFORMATION AND DESCRIPTION OF THE INSTRUMENT

Test Name: Emotional Quotient Inventory 2.0

Date of current review: August 2018

Date of previous review: April 2006

Original test name: BarOn Emotional Quotient Inventory

Authors of the original test: Dr Reuven Bar-On

Authors of the local adaptation: Not given

Local test distributor/publisher: Several distributors in UK

Publisher of the original version of the test: MHS

Date of publication of current revision/edition: October 2011

Date of publication of adaptation for local use: October 2011

Date of publication of original test: 1997

ISBN: 9781854336439

General description of the instrument

The EQ-i 2.0 measures a set of emotional and social skills that influence the way we:

•   perceive and express ourselves,

•   develop and maintain social relationships,

•   cope with challenges, and

•   use emotional information in an effective and meaningful way.

The EQ-i 2.0 model of Emotional Intelligence is comprised of fifteen factors across five categories of functioning, and is based on a model developed by Dr Reuven Bar-On during his dissertation work, his review of the literature on emotional functioning, and from his own clinical practice. The latest version EQ-i 2.0 has been subject to considerable revision based on subsequent research and feedback. The fifteen scales measured by EQ-i 2.0 are called Self-Regard, Emotional Self-Awareness, Assertiveness, Independence, Empathy, Social Responsibility, Interpersonal Relationship, Stress Tolerance, Impulse Control, Reality Testing, Flexibility, Problem Solving, Self-Actualization, Optimism and Happiness. The instrument is computer administered via an on-line portal although a paper version can be used and the responses input to score and generate reports.

There are three reports as follows:

1. The Leadership Report (with two variations, one designed for the 'client' and the other designed for the 'coach'). This structures the feedback around four key leadership dimensions – Authenticity, Coaching, Insight and Innovation.

2. The Workplace Report (also with two variations, one designed for the 'client' and the other designed for the 'coach'). This structures the feedback around the five main factors each comprising of three underlying scales.

3. The Group Report which brings together results from a number of different people and intended for use when working with teams or groups.

The instrument is designed to be suitable for any participant over eighteen years of age and to be used in occupational, educational and psycho-clinical settings. The instrument is expected to take between 20 and 30 minutes to complete. Training in the use of the instrument is required before it can be purchased and accessed. This training is provided by licensed distributors across the world.

Classification

Content domains:

Other: Emotional and social skills

Emotional intelligence - defined as a set of emotional and social skills that influence the way people perceive and express themselves, develop and maintain social relationships, cope with challenges and use emotional information in an effective and meaningful way.

Intended or main area(s) of use:

Advice, guidance and career choice

General health, life and well-being

Work and Occupational

Description of the populations for which the test is intended

Anyone 18 years or older

The EQ-i 2.0 is appropriate for individuals who are 18 years of age and older. It is advised that people with a third to fourth-grade reading level (9-10 years old) can comprehend the EQ-i 2.0 the inventory should not be administered to youths under the age of 18 without thorough consideration of maturity level. It is also not recommended for individuals who are unwilling to answer honestly or who are disoriented or severely impaired.

Number of scales and brief description of the variables) measured by the instrument

15 subscales - Self-Regard, Self-Actualization, Emotional Self-Awareness, Assertiveness, Independence, Empathy, Social Responsibility, Interpersonal Relationship, Stress Tolerance, Impulse Control, Reality Testing, Flexibility, Problem Solving, Self-Actualization, Optimism and Happiness.

The EQ-i 2.0 has 16 scales organised into 5 composites reflecting the different aspects of trait-based emotional intelligence. These are:

Self-Perception Composite:

- Self-Regard subscale: respecting oneself, confidence
- Self-Actualization subscale: pursuit of meaning; self-improvement
- Emotional Self-Awareness subscale: understanding own emotions

Self-Expression Composite:

- Emotional Expression subscale: constructive expression of emotions
- Assertiveness subscale: communicating feelings, beliefs; non-offensive
- Independence subscale: self-directed; free from emotional dependency

Interpersonal Composite:

- Interpersonal Relationships subscale: mutually satisfying relationships
- Empathy subscale: understanding, appreciating how others feel
- Social Responsibility subscale: social consciousness; helpfulness

Decision Making Composite:

- Problem Solving subscale: find solutions when emotions are involved
- Reality Testing subscale: objective; see things as they really are
- Impulse Control subscale: resist or delay impulse to act

Stress Management Composite:

- Flexibility subscale: adapting emotions, thoughts and behaviours
- Stress Tolerance subscale: coping with stressful or difficult situations
- Optimism subscale: positive attitude and outlook on life.

There is also a Well-Being Indicator (Happiness)

There are also four validity scales, which are used to check the accuracy or seriousness of the responses given: omission rate, inconsistency index, positive impression management, and negative impression management.

Response mode

Paper & pencil

Computerised

Demands on the test taker:

Manual capabilities

Irrelevant/not necessary

Handedness

Irrelevant / not necessary

Vision

information missing

Hearing

Irrelevant / not necessary

Command of test language

necessary information given

Reading

Minimum requirement of 3rd or 4th grade reading (p.11 of the manual)

Writing

Irrelevant / not necessary

Items format

Likert scale ratings

- Number of alternatives: 5
- response alternatives (Never/Rarely, Occasionally, Sometimes, Often, Always/Almost Always)

Ipsativity:

- No, multiple choice mixed scale alternatives NOT resulting in ipsative scores

Total number of test items and number of items per scale or subtest

The questionnaire has 133 items distributed as follows:

- Self-Regard subscale: 8 items
- Self-Actualization subscale: 9 items
- Emotional Self-Awareness subscale: 7 items
- Emotional Expression subscale: 8 items
- Assertiveness subscale: 7 items
- Independence subscale: 8 items
- Interpersonal Relationships subscale: 8 items
- Empathy subscale: 9 items
- Social Responsibility subscale: 6 items
- Problem Solving subscale: 8 items
- Reality Testing subscale: 8 items
- Impulse Control subscale: 8 items
- Flexibility subscale: 8 items
- Stress Tolerance subscale: 8 items
- Optimism subscale: 8 items
- Happiness scale: 8 items
- Impression Management scale: 6 items
- Honesty Item: 1 item

Intended mode of use:

- Controlled mode: No direct human supervision of the assessment session is involved but the test is made available only to known test-takers. Internet tests will require test-takers to obtain a link from the test administrator. These often are designed to operate on a one-time-only basis.

Administration mode(s):

Computerised web-based application – unsupervised/self-assessment

Time required for administering the instrument

Preparation: 5 minutes

Administration: 20-40 mins (but a 60-minute window is recommended)

Scoring: 5 minutes (for data input for paper version only)

Analysis: variable

Feedback: variable

Indicate whether different forms of the instrument are available and which form(s) is (are) subject of this review

It is possible to add up to 5 different items identified by publisher.

The same questionnaire is available to be used as part of a 360-feedback process (reviewed separately) and the publishers mention that there is a version called EQ-i Higher Ed but this is also not part of this review.

Measurement and scoring

Scoring procedure for the test:

Computer scoring with direct entry of responses by test taker

Computer scoring with manual entry of responses from the paper response form

Scores:

EQ-i 2.0 standard scores are calculated from raw scores so that each scale has the same average (mean) score of 100 and a standard deviation of 15. This gives an effective range of 70-130 which is displayed on the results graphs.

Specifically, the Total EI score is computed as the sum of all the relevant items on the test (i.e., the 118 items that load onto the Total EI score) and this sum becomes the Total EI raw score. This raw score is then compared to the mean (M) and standard deviation (SD) for the raw score in the normative sample to compute the Standard Score with this formula: Standard Score = (raw score – M)/SD x 15 + 100. Similarly, each Composite and Subscale score is computed as the sum of all the relevant items, and these raw scores are compared to the Means and SDs from the normative sample to compute the Standard Scores.

Lastly, for the EQ-i 2.0 standard scores, 90% confidence intervals are calculated.

Scales used:

Other (please describe): Normative scores with a Mean:100   SD: 15

Score transformation for standard scores:

Normalised – standard scores obtained by use of normalisation look-up table (note: this is done by electronic scoring).

Computer- Generated Reports

Are computer generated reports available with the instrument?

Yes

Name or description of report: **The Workplace Report** focuses on the impact of emotional intelligence at work and offers suggestions for working with colleagues, supervisors, and clients in a variety of coaching, development, and work settings. It has a 'Client' version which contains more narrative interpretation and a Coach version which contains more details concerning the scores and responses to specific items as well as follow-up questions to probe further.	
Media	• Integrated text and graphics
Complexity	• Medium (A mixture of simple descriptions and some configurations of scale scores, and scale interactions)
Report structure	• Scale based – where the report is built around the individual scales.
Sensitivity to context	• Pre-defined context-related versions; number of contexts: 2 contexts (for the client and for the coach)
Clinical-actuarial	• Based on clinical judgment of group of experts
Modifiability	• Limited modification (limited to certain areas, e.g. biodata fields)
Degree of finish	• Publication quality
Transparency	• Clear linkage between constructs, scores and text
Style and tone	• Guidance/suggests hypotheses
Intended recipients	• Qualified test users • Qualified system users • Test takers • Third parties
Do distributors offer a service to modify and/or develop customised computerised reports?	• No

Name or description of report: The Leadership Report focuses on the impact of emotional intelligence for leaders using four key dimensions of leader-ship: authenticity, coaching, insight, and innovation. It contains insights on the leadership and organizational implications of clients' results, as well as strategies for development and potential derailers aimed to help clients reach their leader-ship potential. As an option, the scores can be compared against a group of 220 top leaders. It also has a 'Client' version and a Coach version	
Media	• Integrated text and graphics
Complexity	• Complex (Contains descriptions of patterns and configurations of scale scores, and scale interactions)
Report structure	• **Construct based – where the** report is built around one or more sets of constructs (e.g. in a work set-ting these could be such as team types, leadership styles, or tolerance to stress; in a clinical setting these could be different kinds of psychopathology; etc.) which are linked to the original scale scores.
Sensitivity to context	• Pre-defined context-related versions; number of contexts: 2 contexts (for the client and for the coach)
Clinical-actuarial	• Based on clinical judgment of group of experts
Modifiability	• Limited modification (limited to certain areas, e.g. biodata fields)
Degree of finish	• Publication quality
Transparency	• Clear linkage between constructs, scores and text
Style and tone	• Guidance/suggests hypotheses
Intended recipients	• Qualified test users • Qualified system users • Test takers • Third parties
Do distributors offer a service to modify and/or develop customised computerised reports?	• No – but can select which sections to include

Name or description of report: Group Report	
This report combines the scores of individuals in a manner that will allow for interpretations to be made at the group or team level. An overview of group results is presented to assist with identifying group strengths as well as the areas where the group as a whole could still improve. The organizational implications of a group's emotional intelligence score is presented and strategies for action are recommended to further develop the group's potential.	
Media	• Integrated text and graphics
Complexity	• Medium (A mixture of simple descriptions and some configurations of scale scores, and scale interactions)
Report structure	• Scale based – where the report is built around the individual scales.
Sensitivity to context	• Pre-defined context-related versions; number of contexts: 2 contexts (for the client and for the coach)
Clinical-actuarial	• Based on clinical judgment of group of experts
Modifiability	• Limited modification (limited to certain areas, e.g. biodata fields)
Degree of finish	• Publication quality
Transparency	• Clear linkage between constructs, scores and text
Style and tone	• Guidance/suggests hypotheses
Intended recipients	• Qualified test users • Qualified system users • Test takers • Third parties
Do distributors offer a service to modify and/or develop customised computerised reports?	• No

Supply Conditions and Costs

Documentation provided by the distributor as part of the test package:

User Manual

Technical (psychometric) manual

Supplementary technical information and updates (e.g. local norms, local validation studies etc.)

Methods of publication

Paper

Internet download

Start-up costs:

Certification costs: Users are required to qualify in the EQ-i 2.0 on a 2-day face-to-face training course or the equivalent as distance learning using webinars. Delegates are required to have a Test User Personality (Level B) qualification or has completed a graduate-level courses in tests and measurement at a university or has received equivalent documented training.

Certification is then offered through a network of trainers who typically charge between $1,500 and $2,500 per person for a 2-day programme. There are no set-up or licensing fees for new users. Once certified users open a TAP account through which they have access to the online EQ-i 2.0.

Lastly, users have the option of accessing the EQ-i 2.0 User's Handbook for free as a digital web version or to purchase a paper copy for $150

Recurrent costs: None: Apart from the costs for generating reports (see below) there are no recurrent costs associated with administration. If paper forms are being used these may be downloaded for free from the Talent Assessments Portal . The cost is based on the purchase of tokens priced in US dollars (that are used up when reports are generated) but approximate UK costs are:

Workplace version (for 'client' and 'coach' versions) £56 ex VAT

Leadership version (for 'client' and 'coach' versions) £84 ex VAT

Prices for reports generated by user installed software: As above

Prices for reports generated by postal/fax bureau service: n/a

Prices for reports by internet service: When generating reports through TAP, users purchase tokens which will be deducted from their ac-count with each report. 1 token costs 1 US dollar and are purchased and loaded into users' TAP ac-counts.

EQ-i 2.0 Leadership Report = 90 tokens

EQ-i 2.0 Workplace Report = 60 tokens

EQ-i 2.0 Group Report = 225 tokens

Prices for other bureau services: correcting or developing automatic reports: n/a

Test-related qualifications required by the supplier of the test:

Test-specific accreditation

Professional qualifications required for use of the instrument:

None

Specialist qualification equivalent to EFPA Test User Standard Level 3

Other (indicate): EQ-i 2.0 results must be interpreted by a qualified psychologist, certified individual, or other professional with a master's level course in tests and measurements. Professionals without graduate-level university credits in tests and measurement are required to become accredited users in order to purchase and interpret EQ-i 2.0 results.

EVALUATION OF THE INSTRUMENT

Key to symbols:

[n/a]	This attribute is not applicable to this instrument
0	Not possible to rate as no, or insufficient information is provided
★	Inadequate
★★	Adequate
★★★	Good
★★★★	Excellent

Quality of the explanation of the rationale, the presentation and the information provided

Quality of the explanation of the rationale

Overall rating of the quality of the explanation of the rationale   ★★★

Theoretical foundation of the constructs	★★★
Test development (and/or translation or adaption) procedure	★★★
Thoroughness of the item analyses and item analysis model	★★★
Presentation of content validity	★★★
Summary of relevant research	★★★

Adequacy of documentation available to the user (user and technical manuals, norm supplements, etc.)

Overall adequacy of documentation available to the user (user and technical manuals, norm supplements, etc.)   ★★★

Rationale	★★★
Development	★★
Development of the test through translation/adaption	0
Standardisation	★★★★
Norms	★★★
Reliability	★★★
Construct validity	★★★★
Criterion validity	★★★
Computer generated reports	★★★★

Quality of the procedural instructions provided for the user

Overall adequacy   ★★★★

For test administration	★★★★
For test scoring	★★★
For norming	★★★
For interpretation and reporting	★★★★
For providing feedback and debriefing test takers and others	★★★★
For providing good practice issues on fairness and bias	★★★★
Restrictions on use	★★★★
Software and technical support	0
References and supporting material	★★★★
Quality of the procedural instructions provided for the user	★★★★

Reviewer's comments on the documentation

The EQ-i 2.0 User Manual is clear and well-organized. It gives enough information for a potential user to decide whether or not to use the test and navigates them through the assessment and feedback procedures. There is additional documentation on-line via the EQ-i portal. The online documentation is comprehensive and regularly updated, although not particularly easy to navigate around the information. There is detailed explanation of the various elements, with links to tables of data to illustrate the findings. For example, there is comprehensive discussion about the concept of 'validity' and reference to a variety of academic studies into the issue. The individual findings are generally presented in a straight-forward and comprehensible way, making it clear where they refer specifically to Version EQ-i2.0 of the instrument. There are also some minor inconsistencies regarding the number of items per scale presented in tables A.11 where it states that Interpersonal Relationships has 7 items although there are actually 8 (and there are similar inconsistencies for Problem Solving and Reality Testing).

The instrument is straightforward for the participants to complete, and the system itself is intuitive for the administrator to navigate, both in terms of administering the test and compiling the reports. The instrument is available in a restricted number of other languages and a comprehensive explanation is provided to reinforce the importance of using the instrument in context with other data/information.

The norms are well described with relevant demographics details (gender, age, employment etc.) with effect sizes reported. One aspect that is missing is a description of the context or purpose of the assessments which can have a significant effect on scores (i.e. self-report in evaluation contexts such as selection versus personal development contexts).

The slightly lower ratings for rationale and development may seem surprising since the EQ-i has a long and illustrious history. Its initial conception was from Reuven Bar-On's clinical experience and academic research. Furthermore, the literature review provides a good set of references. However, there is little discussions of the nature of the emotional intelligence (see, for example, Waterhouse, 2010 and similar articles) and since there is still controversy about its nature, breadth and depth the working definition offered is rather broad and general (i.e. EI reflects one's overall well-being and ability to succeed in life). This could be applied to many other psychological constructs (such as psychological capital or personality resources). This makes evaluating how the items were written and what criteria were used for acceptance or rejection rather difficult.

A further issue that is unexplained is the choice of time (a measure of quantity) rather than a measure of quality as the indicator for all of the items (i.e. the scale goes from never, occasionally, sometimes of-ten, always). Time may not always be a good indicator of depth or commitment. An example is item 61 (I contribute to my community) where always seems impossible to endorse. Also, if someone contributes a lot but feels they should do more, they are likely to give a very different answer to someone who does very little but sees it as being more than they really need to. Another example is item 52 (I avoid hurting the feelings of others). Some people may be very conscious about avoiding hurting others because they know it is their tendency. Others may never need to avoid this because they are genuinely kind/empathic.

Quality of the test materials

Quality of the test materials of CBT and WBT

Quality of the design of the software (e.g. robustness in relation to operation when incorrect keys are pressed, internet connections fail etc.)	★★★★
Ease with which the test taker can understand the task	★★★★
Clarity and comprehensiveness of the instructions (including sample items and practice trials) for the test taker, the operation of the software and how to respond if the test is administered by computer	★★★
Ease with which responses or answers can be made by the test taker	★★★
Quality of the design of the user interface	★★★★
Security of the test against unauthorized access to items or to answers	★★★
Quality of the formulation of the items and clarity of graphical content in the case of non-verbal items	★★★
Quality of the materials of CBT and WBT	★★★

Reviewer's comments on quality of the materials

Overall, the quality of the test materials is very good. The test is web-based and can be accessed via personal invitations or via link generated by an administrator. The software seems quite robust, the instructions simple and clear and there are very few elements on the screen, which leaves virtually no room for errors of input. For this reason and because the questionnaire is untimed, the lack of practice questions is probably unnecessary. A minor issue is the small size of the font for the instructions and for some of the items.

Norms

Is the test norm referenced? Yes

Norm referenced interpretation

Overall Adequacy: ★★★

Appropriateness for local use	★★★
Appropriateness for intended applications	★★★
Sample sizes (classical norming)	★★★★
Sample sizes continuous norming	n/a
Procedures used in sample selection	Non-probability sample – quota
Representativeness of the norm sample(s)	★★★
Quality of information provided about minority/protected group differences, effects of age, gender etc.	★★★
How old are the normative studies?	★★★★
Practice effects	n/a

Is the test criterion referenced? No

Reviewer's comments on the norms

The User Manual for the EQ-i 2.0 provides a good description of the main pilot and norming studies based on the North American sample including the process of data collection and norming procedures – with the exception of not defining the context or purpose of the assessments. The normative studies are fresh and all norms are less than 10 years old.

The user would need to be careful when selecting alternative norms such as for student groups where it is unclear. For example, should one use age appropriate norms or should one consider the level of education (which is proved to be a significant predictor of the emotional intelligence in the "Validity" section). It is also unclear how the norms reflect occupational differences, both between types of jobs and the seniority level (for example, in the "Validity" section top management shows significant superiority over the middle and low-level employees).

There are geographical and language appropriate norms (Global norms, UK/Ireland, US/Canada, Australia, South Africa, Sweden and Denmark) which the publishers claim to be updating on a regular basis, so users may well be able to get up-to-date information when required.

Reliability

Overall Adequacy: ★★★

Overall Adequacy	
Data provided about reliability	Only one estimate of standard error of measurement given (for each scale or subscale)
Internal consistency:	
Sample size	★★★★
Kind of coefficients reported	Coefficient alpha or KR-20
Size of coefficients	★★★
Reliability coefficients are reported with samples which…..	…. match the intended test takers
Test related reliability-temporal stability:	
Sample size	★★★
Size of coefficients	★★★
Data provided about test-re-test interval	2 studies: 1) 2-4 weeks apart 2) approx. 8 weeks apart
Reliability coefficients are reported with samples which……	…. match the intended test takers
Equivalence reliability:	
Sample size	n/a
Are the assumptions for parallelism met for the different versions of the test for which equivalence reliability is investigated?	n/a
Size of coefficients	n/a
Reliability coefficients are reported with samples which……….	n/a
IRT based method:	
Sample size	n/a
Kind of coefficients reported	n/a
Size of coefficients (based on the final test length)	n/a
Inter-rater reliability:	
Sample size	n/a
Kind of coefficients reported (select as many as applicable)	n/a
Size of coefficients	n/a

Other methods of reliability estimation:	
Sample size	n/a
Results	n/a

Reviewer's comments on reliability

The manual presents the two most common indices that are used to judge reliability. The first is Cronbach's alpha where a large sample (North American, sample size 4000) and the values for all scales exceed 0.8 except for 2 scales at 0.77. This is normally considered to be excellent when there is a clear rationale that ensures that there is sufficient item diversity to cover the defined domain. However, since the item writing/domain coverage has not been explained, it is not possible to state with certainty that these figures do not represent some measurement redundancy (i.e. bloated specifics).

The other indices presented are test re-test coefficients based on two studies where the time interval was 2-4 weeks (n=204) and approximately 8 weeks (n=101). These confirm that the scales remain highly consistent over time. Since some views of self-reported Emotional Intelligence suggest that the construct is amenable to change (especially as a result of interventions designed to develop it), it would be useful to understand how and why the participants in these studies completed the EQ-i a second time.

Since both indices are based on North American data there are further questions of how these indices apply to EQ-i in different samples, different languages and different cultures but, on balance, it appears that the EQ-i demonstrates solid reliability.

Validity

Overall Adequacy: ★★

Construct validity:	
Design used	➢ Exploratory Factor Analysis ➢ Confirmatory Factor Analysis ➢ Difference between groups ➢ Correlations with other instruments and performance criteria
Do the results of (exploratory or confirmatory) factor analysis support the structure of the test?	★★★
Do the items correlate sufficiently well with the (sub) test score?	0
Is the factor structure invariant across groups and/or is the test free of item-bias (DIF)?	0

Are the differences in mean scores between relevant groups as expected?	★★★
Median and range of the correlations between the test and tests measuring similar constructs	★
Do the correlations with other instruments show good discriminant validity with respect to constructs and the test is not supposed to measure?	0
If a Multi-Trait-Method design is used, do the results support the construct validity of the test (does it really measure what it is supposed to measure and not something else)?	0
Other, e.g. IRT-methodology, (quasi-) experimental designs (describe):	0
Sample sizes	0
Quality of instruments as criteria or markers	★
How are old are validity studies?	0
Construct validity – Overall adequacy	★
Criterion – related validity:	
Type of criterion study or studies (select as many as applicable)	Concurrent
Sample sizes	★★★
Quality of criterion measures	★★
Strength of the relation between test and criteria	★★
Criterion – related validity – overall adequacy	★★
How old are the validity studies	0

Reviewers' comments on validity

The validity studies for the EQ-i 2.0 seem to be relatively recent and cover various aspects of the test validity. The manual provides useful descriptions and data tables helping the user to make informed judgements about how to use the test. The number of studies and methods compares well with what is usually found in test manuals. However, given the breadth and depth of the construct Emotional Intelligence there is a need for more extensive studies. Some of the issues are as follows:

1. The Factor Analyses are thorough and technically competent but the authors conclusions could be challenged. Hence the authors claim that the CFA results provide a good model fit to the model (presumably the 5 major factors and the 15 underlying scales). However, the cut-off for the RMSEA fit index (below 0.10) may be a little high (see, for example, Hooper, Coughlan and Mullen, 2008; they propose 0.06 as a reliable cut-off for RMSEA). Also, the large positive manifold between all the scales raises questions about the rational model and the degree of differentiation implied by the model.

2. Correlations between the EQ-i 2.0 and other instruments deemed to measure the same constructs are not necessarily confirmatory. For example, correlations with MSCEIT are low. The authors suggest that this could be because EQ-i 2.0 is trait-based and MSCEIT is ability-based. However, if a trait obtained for self-report does not correlate with actual competence of that trait its value should be questioned. Of course, MSCEIT may be a poor measure of EI competence, but the explanation provided is unsatisfactory. Similarly, the correlations with the SSI are broadly positive but not clearly differentiated.

3. The correlations with personality (NEO FFI) are in the same direction as found in other studies. How this helps is unclear since the purpose of EI is to identify elements that are beyond personality and more amenable to change and development. Furthermore, it does not address a difficult issue – that personality questionnaires usually claim that there is no right/wrong or good/bad in personality scales. However, there is an underlying value that it is good to be more emotionally intelligent and hence this value gets transmitted to the personality context where it would be deemed positive to be more extravert, open, conscientious, agreeable and less anxious.

4. The results from the criterion studies are a step in the right direction, but there is an over dependence on other self-report instruments (e.g. MLQ to identify transformational leadership styles) where there will be a conflation between the instruments by people who believe they are emotionally intelligent. The hypotheses are unclear or vague and general concepts of success such as job advancement and academic achievement are not clearly related to operational definitions.

5. Reporting low correlations (such as between WGCTA and EQ-i) does show that the constructs are dis-similar, but this can be applied to many things. Also, both EQ-i and WGCTA show positive correlations with academic achievement. Hence the low correlation between EQ-i and WGCTA could do with a little more explanation and hypotheses testing if it is to be presented as validity evidence.

Quality of computer generated reports

Overall adequacy of computer-generated reports: ★★★★

The Workplace Report

Scope or coverage	★★★
Reliability	★★★★
Relevance or validity	★★★
Fairness, or freedom from systematic bias	★★★★
Acceptability	★★★★
Length	★★★★

The Leadership Report

Scope or coverage	★★★
Reliability	★★★★
Relevance or validity	★★★
Fairness, or freedom from systematic bias	★★★★
Acceptability	★★★★
Length	★★★★

Group Report

Scope or coverage	★★★
Reliability	★★★★
Relevance or validity	★★★
Fairness, or freedom from systematic bias	★★★★
Acceptability	★★★★
Length	★★★★

Reviewers' comments on computer generated reports:

The computer-generated reports are well-organized and easy-to-use. The language is clear and precise. Whilst the whole report may be a little long for typical use, there is a facility to exclude or include various sections and such tailoring may prove very useful. There are good narrative explanations of the characteristics/behaviours being measured with a useful "Balancing your EI" section describing how some of the interactions between scales. The information is presented both in textual and graphical forms, which effectively complement one another.

Nevertheless, the reports are unlikely to stand alone and be decipherable by an untrained participant without input from a qualified person. The Coach reports give useful guidance for the discussion of the results with the client: the scales description, the verbal and graphic representation of the test results, questions for further discussion. The authors do provide training and recommend that a qualified person manages the interpretation especially for concepts that the lay person is unlikely to understand (such as the norm group and the rationale behind the model).

Developmental guidance is included in the reports, for example in the form of questions, worksheets and an action plan. The Leadership report ends with a summary of the characteristics and 'common trends' across generations. This runs the risk of appearing to stereotype and 'label' individuals and this may be over-interpretation and potentially unhelpful.

Final Evaluation

Evaluative report of the test:

The EQ-i 2.0 represents a significant body of work that attempts to measure the complex and sometimes ill-defined domain known as emotional intelligence. From its early development in 1998 it has been updated and improved. However, the emotional intelligence domain is so vast (covering personal, social and emotional development of the broad adult population) that no single instrument will cover all of the territory and the many facets will take a long time to validate. However, the EQ-i represents one of the more serious and impressive instruments in this area. It is one of the most comprehensive and it demonstrates levels of reliability that meet traditional psychometric expectations. The normative data available is growing and there is some evidence of its validity which is very promising. The Manual provides a good description of the test administration and feedback processes. Computerised user reports are well-structured, use clear and friendly language and combine textual feedback and colourful graphics.

The questions that need further attention are summarised below.

1. The definition of the emotional intelligence could be tightened

2. The structure may be rationally useful but may involve measurement redundancy

3. Clearer hypotheses would be useful for evaluating its validity

4. Better description of the norms beyond the North American sample for those using it in different languages or with different groups

Conclusions:

EQ-i 2.0 is an interesting and useful instrument for those who embrace the concept of emotional intelligence. It appears suitable for the broad adult population – certainly in North America and it appears to be suitable in other countries and cultures although this does require further evidence. It is primarily a coaching and career guidance tool and its use in an evaluation or similar high stakes context should be approached with great caution. There is insufficient evidence to confirm its robustness against low self-awareness, self-delusion and impression management.

The reports provide users with well-structured information on the current level of their emotional intelligence – as they believe it to be – and helps them with recommendations for further development.

Recommendations:

> ➢ Suitable for use in the area(s) of application defined by the distributor, by test users who meet the distributor's specific qualification requirements (at least EFPA User Qualification Level 2)

Bibliography:

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. European Journal of Psychological Assessment, 12, 62–71.

Bartram, D. (2002a). EFPA Review Model for the description and evaluation of psychological instruments: Version 3.2. Evaluation Form. Brussels: EFPA Standing Committee on Tests and Testing (September, 2002).

Bartram, D. (2002b). EFPA Review Model for the description and evaluation of psychological instruments: Version 3.2. Notes for Reviewers. Brussels: EFPA Standing Committee on Tests and Testing (September, 2002).

Bartram, D., & Hambleton, R. K. (Eds.) (2006). Computer-based testing and the Internet. Chichester, UK: Wiley and Sons.

Bartram, D., Lindley, P. A., & Foster, J. M. (1990). A review of psychometric tests for assessment in vocational training. Sheffield, UK: The Training Agency.

Bartram, D., Lindley, P. A., & Foster, J. M. (1992). Review of psychometric tests for assessment in vocational training. BPS Books: Leicester.

Bechger, T., Hemker, B., & Maris, G. (2009). Over het gebruik van continue normering[On the use of continuous norming]. Arnhem, The Netherlands: Cito.

Bennett, R. E. (2006). Inexorable and inevitable: The continuing story of technology and assessment. In D. Bartram & R. K. Hambleton (Eds.), Computer-based testing and the Internet (pp. 201-217). Chichester, UK: Wiley and Sons.

Brennan, R. L. (Ed.) (2006). Educational measurement. Westport, CT: ACE/Praeger.

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.

Downing, S. M., & Haladyna, T. M. (Eds.) (2006). Handbook of test development. Hillsdale, NJ: Erlbaum.

Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), Educational measurement (pp. 471-515).Westport, CT: ACE/Praeger.

Drenth, P. J. D., &Sijtsma, K. (2006). Testtheorie. Inleiding in de theorie van de psychologische test enzijntoepassing-en(4e herzienedruk) [Test theory. Introduction in the

theory and application of psychological tests (4th revised ed.)]. Houten, The Netherlands: Bohn Stafleu van Loghum.

Embretson, S. E. (Ed.) (2010). Measuring psychological constructs. Advances in model-based approaches. Washington, D. C.: American Psychological Association.

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

Evers, A. (2001a). Improving test quality in the Netherlands: Results of 18 years of test ratings. International Journal of Testing, 1, 137–153.

Evers, A. (2001b). The revised Dutch rating system for test quality. International Journal of Testing, 1, 155–182.

Evers, A., Braak, M., Frima, R., & van Vliet-Mulder, J. C. (2009-2012). Documentatie van Tests enTestresearch in Nederland [Documentation of Tests and Testresearch in The Netherlands]. Amsterdam: Boom test uitgevers.

Evers, A., Lucassen, W., Meijer, R., &Sijtsma, K. (2010). COTAN Beoordelingssysteemvoor de Kwaliteit van Tests (ge-heelherzieneversie; gewijzigdeherdruk) [COTAN Rating system for test quality (completely revised edition; re-vised reprint)]. Amsterdam: NIP.

Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., et al. (2012). Testing practices in the 21st Century: Developments and European psychologists' opinions. European Psychologist, in press.

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure and results. International Journal of Testing, 10, 295-317.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15, 309-334.

Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. Applied Psychological Measurement, 24, 355–366.

Hambleton, R. K., Merenda, P. F., &Spielberger, C. D. (Eds.) (2005). Adapting educational and psychological tests for cross-cultural assessment. Mahwah, NJ: Erlbaum.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. American Psychologist, 58, 78-80.

International Test Commission. (2005). International Guidelines on Computer-Based and Internet Delivered Testing. Bruxelles, Belgium: Author.

Kersting, M. (2008). DIN Screen, Version 2. LeitfadenzurKontrolle und Optimierung der Qualität von Verfahren und derenEinsatzbeiberuflichenEignungsbeurteilungen [DIN Screen, Version 2. Guide line for monitoring and opti-mizing the quality of instruments and their application in proficiency assessment procedures.]. In M. Kersting. Qualitätssicherung in der

Diagnostik und Personalauswahl - der DIN Ansatz (S. 141-210) [Guaranteeing quality in diagnostics and personnel selection (p. 141-210)]. Göttingen: Hogrefe.

Lindley, P. A. (2009). Reviewing translated and adapted tests: Notes and checklist for reviewers:5 May 2009. Leicester, UK: British Psychological Society. Retrieved from http://www.efpa. eu/professional-development/tests-and-testing.

Lindley, P.A. (2009, July). Using EFPA Criteria as a common standard to review tests and instruments in different countries. In D.Bartram (Chair), National approaches to test quality assurance. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.

Lindley, P., Bartram, D., & Kennedy, N. (2004). EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.3. Leicester, UK: British Psychological Society (November, 2004).

Lindley, P., Bartram, D., & Kennedy, N. (2005). EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.41. Brussels: EFPA Standing Committee on Tests and Testing (August, 2005).

Lindley, P., Bartram, D., & Kennedy, N. (2008). EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.42. Brussels: EFPA Standing Committee on Tests and Testing (September, 2008).

Lindley, P. A. (Senior Editor), Cooper, J., Robertson, I., Smith, M., & Waters, S. (Consulting Editors). (2001). Review of personality assessment instruments (Level B) for use in occupational settings. 2nd Edition. Leicester, UK: BPS Books.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist, 56, 128-165.

Mokken, R. J. (1971). A theory and procedure of scale analysis. The Hague: Mouton.

Moosbrugger, H., Kelava, A., Hagemeister, C., Kersting, M., Lang, F., Reimann, G., et al. (2009, July).The German Test Review System (TBS-TK) and first experiences. In D. Bartram (Chair), National approaches to test quality assurance. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.

Moreno, R., Martínez, R. J., &Muñiz, J. (2006). New guidelines for developing multiple-choice ítems. Methodology, 2, 65-72.

Muñiz, J., & Bartram, D. (2007). Improving international tests and testing. European Psychologist, 12, 206-219.

Nielsen, S. L. (2009, July). Test certification through DNV in Norway. In D. Bartram (Chair), National approaches to test quality assurance. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rded.). New York: McGraw-Hill.

Parshall, C. G., Spray, J. A., Davey, T., &Kalohn, J. (2001). Practical Considerations in Computer-based Testing. New York: Springer Verlag.

Prieto, G., &Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizadosenEspaña [Amodel for the evaluation of test quality in Spain]. Papeles del Psicólogo, 77, 65–71.

Reise, S. P., &Havilund, M. G. (2005). Item response theory and the measurement of clinical change. Journal of Per-sonality Measurement, 84, 228-238.

Tideman, E. (2007). Psychological tests and testing in Sweden. Testing International, 17(June), 5–7.

Schneider, R. J., & Hough, L. M. (1995). Personality and industrial/organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), International Review of Industrial and Organizational Psychology, 10, 75-129.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. Statistical Methods in Medical Research, 7, 301-317.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. Science, 240, 1285-1293.

Testkuratorium. (2009). TBS-TK. Testbeurteilungssystem des Testkuratoriums der FöderationDeutscherPsychologen-vereinigungen. RevidierteFassungvom 09. September 2009 [TBS-TK. Test review system of the board of testing of the Federation of German psychologists' associations]. Report Psychologie, 34, 470-478.

Van de Vijver, F. J. R., &Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Ham-bleton, P. F. Merenda, & C. D. Spielberger (Eds.), Adapting educational and psychological tests for cross-cultural assessment. Mahwah, NJ: Erlbaum.

Van der Linden, W. J., &Glas, C. A. W. (Eds.) (2010). Elements of adaptive testing. London: Springer.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Erlbaum.

Ziegler, M., MacCann, C., & Roberts, R. (Eds.) (2011). New perspectives on faking in personality assessment. Oxford, UK: Oxford University Press.