The British Psychological Society
Psychological Testing Centre

Test Review

Hogan Personality Inventory (HPI)

Test Review of Hogan Personality Inventory

Reviewers: Iain Coyne & Daniel Hinton

Consultant Editor: Louisa Tate

Senior Editor: Charles Eyre

GENERAL INFORMATION AND DESCRIPTION OF THE INSTRUMENT

Test Name: Hogan Personality Inventory

Date of current review: 2018

Date of previous review: 2009

Original test name: Hogan Personality Inventory

Authors of the original test: Robert Hogan and Joyce Hogan

Authors of the local adaptation: Robert Hogan and Joyce Hogan

Local test distributor/publisher: Psychological Consultancy Ltd; 3 Minute Mile; Advanced People Strategies; Mentis.

Publisher of the original version of the test: Hogan Assessments

Date of publication of current revision/edition: 2007

Date of publication of adaptation for local use: 2007

Date of publication of original test:1992

ISBN: 9781854336231

General description of the instrument

The Hogan Personality Inventory (HPI) is a trait measure of personality, based upon the California Psychological Inventory. Originally published in 1992, the current version of the tool was released in 2007. It is a personality instrument designed to assess normal personality in adults - predominantly in work and occupational settings for selection and development purposes, although it can be used in other non-clinical contexts, such as for research purposes and in counselling settings. It is based on Socioanalytic Theory in which people desire to 'get along' with and 'get ahead' of other group members. This approach suggests that individuals respond to questionnaires as a way of informing others how they wish to be regarded.

The HPI is almost exclusively administered online in controlled mode (requiring a username and password), but it can be administered in pencil and paper format in exceptional circumstances. In these cases, it is then scored electronically.

The HPI comprises 206 true/false personality statements. Based on responses to these statements, test takers are scored on seven primary scales (Adjustment, Ambition, Sociability, Interpersonal Sensitivity, Prudence, Inquisitive and Learning Approach) that map the Five-Factor model, which are further subdivided into 41 Homogeneous Item Clusters (HICs), which represent facets of these scales. Additionally, scores on six HPI Occupational Scales are generated for use with specific occupational groups such as managerial, sales and clerical staff. The HPI also contains a validity scale to detect careless responding.

The seven primary scales are:

- Adjustment – calm and self-accepting versus self-critical or tense
- Ambition – socially self-confident, leader-like, competitive and energetic
- Sociability – extent a person needs to or enjoys interacting with others
- Interpersonal sensitivity – extent a person is seen as perceptive, tactful and socially sensitive
- Prudence – the degree an individual appears conscientious, conforming and dependable
- Inquisitive – extent an individual is seen as bright, creative and interested in intellectual matters
- Learning approach – the extent a person enjoys academic activities and values educational achievement

The six occupational scales are:

- Service Orientation – people who treat customers/colleagues in a courteous and helpful manner
- Stress Tolerance – the extent an individual can easily handle stress, pressure and heavy workloads
- Reliability – willingly follow rules and respect corporate values
- Clerical Potential – those with a talent for clerical work and administrative responsibilities
- Sales Potential – a talent for sales
- Managerial Potential – building and maintaining effective teams

The HPI is available in a range of languages, and has norm groups to represent a variety of local populations. There is a revised 2005 US norm group based on a sample of 156,614 US working adults as well as a UK norm group comprising 64,768 working adults' data collected between August 2004 and September 2014. The HPI also has a global norm, which aggregates data from the 46 different language versions of the HPI, and includes 171,132 cases.

HPI scale scores are normalised using percentiles and can be used to generate a wide range of reports for use in different occupational contexts. Reports are available that focus on individuals and teams, entry level staff, managers/leaders (including first time managers), hourly employees, sales, and safety-critical roles. The 'Insight' report provides percentile scores for each primary and occupational scale as well as a graphical presentation of the HICs which comprise the primary/occupational scale. The test-taker is told that these HICs should be interpreted by a qualified Hogan coach. The report is developmental in its approach, providing a description of the scale, short descriptors of the test-taker's personality in relation to a scale, and a series of discussion points which allow further exploration. The 'Potential' report interprets scale scores in relation to strengths and competencies for leadership. Again, this report is descriptive and developmental in nature, providing percentile scores, behavioural implications, leadership implications, competency analysis and developmental recommendations in relation to the competency analysis.

Test users do not need a specific professional qualification to use the HPI, but Hogan Assessment Systems require users to complete HPI training and certification.

Classification

Content domains:

Personality – Trait

Intended or main area(s) of use

Work and Occupational

Description of the populations for which the test is intended

Normal population of adults

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

The HPI comprises 41 homogeneous item clusters (HICs) which are used to derive scores on seven primary scales:

Adjustment – calm and self-accepting versus self-critical or tense

Ambition – socially self-confident, leader-like, competitive and energetic

Sociability – extent a person needs to or enjoys interacting with others

Interpersonal sensitivity (previously known as either Agreeability or Likeability) – extent a person is seen as perceptive, tactful and socially sensitive

Prudence – the degree an individual appears conscientious, conforming and dependable

Inquisitive (previously Intellectance)– extent an individual is seen as bright, creative and interested in intellectual matters

Learning approach (previously School Success) – the extent a person enjoys academic activities

There is also a validity scale which measures careless responding.

Additionally, scores are provided for six occupational scales:

Service Orientation – people who treat customers/colleagues with courtesy

Stress Tolerance – the extent an individual can easily handle stress, pressure and heavy workloads

Reliability – willingly follow rules and respect corporate values

Clerical Potential – those with a talent for clerical work and administrative responsibilities

Sales Potential – a talent for sales

Managerial Potential – building and maintaining effective teams

Response mode

- Paper & pencil
- Computerised

Demands on the test taker:

Manual capabilities

Irrelevant/not necessary

Handedness

Irrelevant / not necessary

Vision

necessary information given

Hearing

Irrelevant / not necessary

Command of test language

necessary information given

Reading

necessary information given

Writing

Irrelevant / not necessary

Items format

Likert scale ratings

Number of alternatives: 2 (True / False) **Ipsativity**:

- Not relevant

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

Total number of items equals 206:

Adjustment (37)

Ambition (29)

Sociability (24)

Interpersonal Sensitivity (22)

Prudence (31)

Inquisitive (25)

Learning Approach (14)

Validity (14)

Adjustment and Interpersonal Sensitivity share one item and so do Prudence and Validity.

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 logon username and password. 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:

1 minute

Administration:

Approx. 15 minutes

Scoring:

Automated, so generally instant (up to 5 minutes)

Analysis:

Automated, so generally instant (up to 5 minutes)

Feedback:

90 minutes

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

No other forms are currently available

Measurement and scoring

Scoring procedure for the test:

Computer scoring with direct entry of responses by test taker

Computer scoring by Optical Mark Reader entry of responses from the paper response form

Scores:

Items are keyed to a specific scale and HIC and are unit-weighted.

Test-takers are required to respond to at least two-thirds of the items in a scale and HIC. Mean substitution is used to estimate missing item values if needed.

Scales used

Percentile Based Scores - Centiles

Score transformation for standard scores:

Normalised – standard scores obtained by use of normalisation look-up table

Computer- Generated Reports

Are computer generated reports available with the instrument?

Yes

Name or description of report: Insight	
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: ....
Clinical-actuarial	• Based on empirical/actuarial relationships

Modifiability	• Limited modification
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?	• Yes

Name or description of report: Potential (part of the Hogan LEAD series)	
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: ....
Clinical-actuarial	• Based on empirical/actuarial relationships
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?	• Yes

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.)

Books and articles of related interest

Methods of publication:

Paper

Internet download

Start – up costs:

Hogan Assessment Systems require Hogan training and certification, which involves a 2-day workshop. Cost = £1500.

Additional certification workshops are available for users who would like a deeper understanding of the assessments and want more experience providing feedback.

Recurrent costs:

n/a

Prices for reports generated by user installed software:

n/a

Prices for reports generated by postal/fax bureau service:

n/a

Prices for reports by internet service:

£30-200 per report (price varies according to report type)

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 (Hogan Assessment Systems requires Hogan training and certification)

Professional qualifications required for use of the instrument:

None

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	★★★★
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	
References and supporting material	
Quality of the procedural instructions provided for the user	

Reviewer's comments on the documentation

Overall, the quality of the information provided to the user is excellent. There is extensive information on the rationale, theoretical positioning and research chronology of the HPI. The generic test manual provides details on the initial development and conceptualisation of the HPI, through to later developments and updates. It offers the reader information on reliability, validity and norms generated within the US. The publisher provides reviewers with additional material on validity, reports, the certification training programme and UK psychometric data. All of the material is informative and in the main provided sufficient detail to understand the scales and its properties. As well as offering information on the development and psychometrics of the HPI, the manual provides some guidance on administering and interpreting the HPI.

While the authors detail how items were created and methods used to reduce item numbers, there is little data to indicate the item analysis approach and results. Some data is provided for the HICs, but not at the item level. More information on the rules used to remove and/or accept items would be helpful.

Additionally, it was not clear in the supplied documentation how the scale scoring actually works via the HICs. On this point, the publishers were able to clarify that each raw HIC score is the sum of its component items, and each raw scale score is the sum of its component HICs. The publishers supply a range of sample reports via their website which demonstrate how the scales are utilised within the wide variety of reports available to the user.

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

The materials presented online are of a good quality. The interface is minimalist in appearance, but as a result the test is easy to follow and complete online. Practice items are not presented, but the instructions are sufficiently clear to be understood and for the items to be responded to without these.

Items are generally appropriate although there were some double-negative items which required a little bit of thought in terms of understanding the correct way to respond. Also, some items refer back to childhood and school which may be influenced too much by false memories rather than presenting an image as suggested by Socioanalytic Theory. The authors have attempted to minimise the implications of the latter problem, through evaluating the relationship between scale / subscale scores and third party ratings using adjective checklists (see validity section).

The web pages appear to be well coded and free from bugs. Though the manual contains an FAQ covering potential issues that might arise (such as loss of internet connection), none of these would be likely to jeopardise the integrity of the measure. Whilst difficult to test fully, the system itself appears secure, and is inaccessible to unauthorised persons. The system is programmed in such a way that results cannot be obtained through hacking of the online system, being processed and held on Hogan's servers.

Norms

Is the test norm referenced? Yes

If yes, please complete the section below:

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 – convenience
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 UK HPI normative sample comprises 64,768 working adults' data collected between August 2004 and September 2014. The sample derives from a convenience sampling approach using existing datasets within the HPI database. While there are potential biases with non-probability and non-representative sampling for norms, this approach is increasingly common within personality testing in occupational settings. The UK norm sample is generally well-defined with data available on gender, age, ethnicity, job family and reason for assessment. Age data indicates approximately 70% of the sample are in the range of under 30 to 49, with 10% between 50-59 and 1% above 60. More males (58.2%) than female (35.8%) are represented as are Whites (60.3%) as compared to other ethnic groups. However, given the total sample size of the norm group, actual numbers within these categories are sufficiently large for appropriate analyses to commence on possible differences in score performance (e.g. only 4.9% of the sample are Asian or Asian British, yet this equates to 3,201 people).

More of a limitation is the data on job family and reason for assessment where in both cases 90% of the sample did not report data on these demographics. This restricts the ability to judge fully the representativeness of the sample and group difference on these demographic parameters. However, extensive analyses have been conducted to examine group differences on the HPI scales. Both Procrustes analysis and DIF detection using the logistic regression method demonstrated that the HPI scales function equivalently for different groups.

The authors have specified that each norm is updated approximately every five years.

Reliability

Overall Adequacy: ★★★

Overall Adequacy	
Data provided about reliability	➢ One reliability coefficient given in technical manual (for each scale or subscale) *(McDonald's Omega co-efficients were also made available through an addendum)* ➢ 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	The interval is: 2 days to 2.58 years (mean = 0.79 years, SD = 0.59 years)
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	★★★

Kind of coefficients reported	IRT derived reliability based on Dimitrov, 2003
Size of coefficients (based on the final test length)	★★★
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 internal consistency and test-retest reliability are examined in two datasets (UK and US samples). This review is based on the data on the UK sample rather than the US data.

UK HPI reliability data is generally based on the norm sample of 64,768 working adults. Using this data estimates are provided for Cronbach's Alpha, mean-corrected item-total correlation, IRT-derived reliability and two measures of SEM (one applied to an estimated true score and one to the observed score). Alpha estimates range from 0.66 for Interpersonal Sensitivity to 0.86 for both Adjustment and Ambition (median = 0.77), whereas IRT-derived reliability estimates range from 0.78 to 0.93 (median = 0.88) for the same scales. These are all adequate with the exception of the Cronbach's Alpha estimate for Interpersonal Sensitivity.

In addition to the standard technical manual, the publishers also provided (post-review) an addendum which included reliability analysis using McDonald's Omega. This alternative analysis has produced somewhat improved reliability coefficients for the scales, ranging from 0.74 (Interpersonal Sensitivity) to 0.88 (Adjustment).

Estimates of test-retest reliability are also provided for a smaller sample of 485 people who took the HPI assessment twice over a range of test intervals from 2 days to 2.58 years. Data shows test-retest values from 0.51 (Interpersonal Sensitivity) to 0.77 (Inquisitive), with a median of 0.69. While this indicates acceptable test-retest reliability for all but Interpersonal Sensitivity, the wide range of retest intervals somewhat clouds interpretation of the analyses here. While demographic information relating to those included within the test-retest reliability study was supplied by the publisher, this did contain a relatively large proportion (>35%) of missing data. s.

UK reliability data was initially provided only for the seven Primary scales, although further UK reliability data was subsequently provided by the publishers in relation to the Occupational Scales and the HICs, This additional information is due to be included in a future edition of the Technical Manual, and covers both internal consistency (McDonald's Omega), and test-retest

reliability. For the Occupational scales, the majority of alpha coefficients were good, ranging from 0.63 to 0.95, Omega coefficients ranging from 0.67-0.87, and test re-test coefficients ranging from 0.57 to 0.69. For the HICs, coefficients were lower, given the relatively small number of items within each (3-6). Omega coefficients ranged from 0.31 – 0.87, and test-retest coefficients ranged from 0.33-0.73.

Validity

Overall Adequacy: 

Construct validity:	
Design used	➢ Exploratory Factor Analysis ➢ Confirmatory Factor Analysis ➢ (Corrected) item-test correlations ➢ Testing for invariance of structure and differential item functioning across groups ➢ 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?	
Is the factor structure invariant across groups and/or is the test free of item-bias (DIF)?	
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?	
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)?	n/a
Other, e.g. IRT-methodology, (quasi-) experimental designs (describe):	n/a
Sample sizes	

Quality of instruments as criteria or markers	★★★★
How are old are validity studies?	Number of years: 15+
Construct validity – Overall adequacy	★★★
Criterion – related validity:	
Type of criterion study or studies	Predictive Concurrent Post-dictive
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	Number of years: 2 – 23 years old for UK studies

Reviewers' comments on validity

Overall, the exploration of validity for the HPI is very thorough. Validity evidence is given via an excellent range of good quality studies, for the purposes of establishing construct and criterion-related validity.

The HPI has a robust approach to test-translation and adaptation, with information provided on the methodology detailed within Technical Report 1. This illustrates the strong relationship between the US and UK norms samples and hence permits the use of US data to help establish evidence of validity.

Construct validity:

A 1992 exploratory factor analysis study on 2,500 employed individuals (no further data on this sample is provided) using HICs as the input variables indicates a favourable seven-factor model. Generally, loadings for HICs are loaded onto only one of the Primary scales, although there are some cross-loadings seen for some HICs. More recently, a 2007 confirmatory factor analysis using a sample of 156,614 individuals from the US-norm group indicated a good fit for the seven-factor model (RMSEA = 0.059), with factor loadings in all but one case above 0.3.

Item-scale correlations reported on the UK norm sample showed corrected values ranging from 0.23 for Prudence to 0.38 for Ambition, with a mean correlation of 0.32.

Correlations with other personality assessments are reported with a number of large US samples but smaller UK samples (N=54 to 65). The dates for the US data are mid to late 1990s; with the exception of a 1997 study using the Prevue ICES, dates are not provided for

the UK studies. Evidence of construct validity with these data is mixed. There are some good scale correlations, but also some low correlations.

Median correlations for Big-5 markers are provided in the manual for each Primary scale across four different studies using different personality tests. They range from 0.30 for Learning Approach with Openness to 0.69 for Adjustment with Neuroticism. Where scales directly map (e.g. Sociability with Extraversion, Adjustment with Neuroticism) there are good relationships (0.58 and 0.69 respectively). However, there are inadequate relationships between Prudence and Conscientiousness (0.47) and between Interpersonal Sensitivity and Agreeableness (0.50). Furthermore, with the exception of one of the studies, samples all consisted of students from the US and Spain. Therefore, a further set of UK studies would be desirable.

Discriminant validity is demonstrated by the pattern of correlations between the HPI scales and unrelated scales from other personality measures. These correlation coefficients are, for the most part, small and non-significant.

Additionally, there is a large amount of analysis on group differences (age, gender, ethnicity) which generally shows low to moderate effect sizes. DIF analyses were conducted for each Primary scale across age, gender and ethnic group and no DIF was observed between groups for any of the scales.

Criterion validity:

There are an impressive number of criterion-related validity studies provided in the technical manual, the separate chapter on validity and supplementary psychometric information. Since the HPI is an established tool, this has allowed the publishers to conduct meta-analysis on previous HPI validity studies, as well as presenting findings from ongoing validity studies. While the most recent study supplied in the main review materials was conducted in 2016, the publishers state that they are continually running criterion validity studies on the HPI, including 10 studies under completion in 2018.

Primary scales, Occupational scales and HICs are correlated with a wide range of work-based criteria including ratings of performance and more objective performance data. Samples used range in size from quite small to relatively large, but the meta-analytic approach used by the authors combines together data into large sample sizes for analysis. Criterion validation data reported is mostly with US samples, but there are a good number of UK-specific research studies using correlational data and group differences. Furthermore, samples reflect a broad section of job levels and industry sectors.

Given the amount of data available it is challenging to provide an accurate estimate of the mean validity for criterion validity. The meta-analytic studies, synthetic validity and more focused predictor-criterion research generally show stronger coefficients. Indeed, some of these are quite large (above 0.5), which is impressive for a personality instrument.

Quality of computer generated reports

Overall adequacy of computer generated reports: 

Report 1: Insight

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

Report 2: Potential

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

Reviewers' comments on computer generated reports:

Overall, the computer-generated reports available for the HPI are of a high quality. There are a wide range of reports available, and – while examples of all of them were not submitted for review – the reports submitted demonstrated that they are fit for purpose. Though the presentation and depth of the reports can be tailored based on intended use, all accurately reflect the scope of the scales measured by the HPI.

The content of the reports is consistent with the scale scores obtained, and the performance implications based on these scores, both positive and negative, are consistent with the constructs they purport to measure. Advice is provided for certain trait score interpretations to check for interactions with other scales if appropriate. The personality profiles presented within the reports have been demonstrated to be consistent with how test takers are viewed by others who know them.

Reports aim to ensure fairness by using gender-neutral text. They are available in 25 languages to ensure that they can be used by a wide range of test takers.

Two reports were submitted for the review – Insight and Potential. Both present percentile scores on each Primary and Occupational scale. Insight interprets the percentile rank in relation to categories of low (percentiles 0-25), below average (26-50), above average (51-75) and high (76). This report will be acceptable to test users and test-takers because it is simple to interpret as well as providing a series of discussion points which a test-taker could use as self-reflection and a test user to adopt to facilitate discussion. Unlike the more general nature of Insight, the Potential report is very much geared to leadership and the mapping of a test-takers score on each scale to leadership behaviours and competencies. This will make it attractive to individuals wishing to use the HPI in leadership development or even promotion decisions.

One query concerning the report content is the apparent inconsistency between the interpretation of the Adjustment score and Stress Tolerance score for the above average scoring example. The Adjustment text suggests the individual would tend to cope well in stressful situations; whereas the Stress Tolerance text implies the person is below average in his/her ability to deal with stress. The publisher has confirmed that this is a result of the dummy data used to create a simulated profile and unlikely to occur in practice.

Final Evaluation

Evaluative report of the test:

The Hogan Personality Inventory (HPI) is a well-established personality measure, being one of the most well-known personality assessments globally. It is focused on measuring normal personality in adults, with the majority of evidence pointing to its use in work and occupational contexts. It is a simple inventory to complete with test-takers being asked to respond True or False to 206 items. The online system and computer-generated reports are simple and appear very easy to use. The system can be used to generate a wide range of reports based on a test taker's HPI scale scores, many of which come with a large number of language options. Hogan are also continually working to develop new translations and localisations of the tool for use in other countries. Reports generated are developmental and easy to understand for test-takers and users (especially as percentile scores are condensed into categorical statements of low, below average, above average and high).

There is a large amount of information provided by the developer/publisher, which is comprehensive and identifies the majority of the necessary information needed for a test user to evaluate the utility of the inventory. Hogan do not require any formal psychometric testing qualifications to purchase and use their tools, but require that users have completed their own Hogan certification training. This training prepares users well to administer and feedback the tool, and includes in depth guidance for doing so.

While there is commonality with the Five-Factor Model of personality, the HPI is based more on the California Psychological Inventory (CPI) and adopts a socioanalytic approach in which individuals respond to questionnaires as a way of informing others how they wish to be regarded.

The inventory adopts a seven-factor model of personality (Adjustment, Ambition, Sociability, Interpersonal Sensitivity, Prudence, Inquisitive and Learning Approach) for which there is exploratory and confirmatory factor analysis evidence to support this structure. There are also

additional Occupational scales that are focused more specifically at the work-related constructs of Service Orientation, Stress Tolerance, Reliability, Clerical Potential, Sales Potential and Managerial Potential. On request, the publishers provided further data around the Construct Validity of the Occupational scales, which cited a number of correlational studies that were generally supportive of the Occupational scales. However, all of these studies were over 20 years old, and this is an area in which the evidence base could be further expanded in the future, particularly in relation to the 'Potential' scales (Sales, Clerical and Managerial).

The norm data is very strong with norms generated for a US sample as well as new norms specifically with a UK context. The UK sample, while collated via opportunity sampling, is very large (N=64,768) with data provided on gender, age, ethnicity, job level and type of study. The latter two demographics are limited in range with the majority of the sample (90%) having no recorded data on these criteria. Evidence of low effect sizes for demographic differences in addition to the lack of DIF across Primary scales indicates support for the use of the HPI across different demographic groups. The publishers also provided DIF analysis for age, gender and ethnicity for the Occupational scales on request, and these will be included in the openly published technical materials in future. The analyses or the Occupational scales again showed relatively low effect sizes.

Reliability evidence based on the UK norm group (and a separate sample of 485 UK adults for the test-retest study) is adequate. Data is only provided for both the Primary scalesand Occupational scales. The test-retest interval range is very large, with intervals of up to 8 years, so it would be prudent to retest test-takers substantially before this, as coefficients decline markedly over this length of time.

Evidence of criterion validity is extensive and in the main positive. There are a large number of studies cited (within the US and UK), using a range of different criteria, different samples and different methodologies (e.g. correlations, meta-analysis and synthetic validity). Construct validity data shows some support, especially internal construct validity via exploratory and confirmatory factor analysis. However, the relationship between the HPI and other personality scales tends to be adequate at best. This may be due in part to the fact that the HPI does not map exactly on to the Five-Factor model. The Sociability and Adjustment scales show good relationships with big-5 markers; whereas the Prudence and Interpersonal Sensitivity scales less so. Given the dates of these studies and the smaller samples seen in the UK, we would encourage the authors to design further construct validity studies of the HPI with UK samples.

Conclusions:

The 206-item Hogan Personality Inventory is a useful tool for use with working adult samples within the UK. Evidence for the seven Primary scale structure and the prediction of job performance is supportive. More UK-based data should be collected for the Occupational scales to support their validity – especially as these scales are included in the reports. The HPI is a simple measure to administer via the controlled online testing format and produces reports instantaneously. These reports provide a basis on which a trained assessor can explore an individual's personality as it applies to a working context.

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 en zijn toepassingen* (4e herziene druk) [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 en Testresearch in Nederland* [Documentation of Tests and Testresearch in The Netherlands]. Amsterdam: Boom test uitgevers.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de Kwaliteit van Tests (geheel herziene versie; gewijzigde herdruk)* [COTAN Rating system for test quality (completely revised edition; revised 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. Leitfaden zur Kontrolle und Optimierung der Qualität von Verfahren und deren Einsatz bei beruflichen Eignungsbeurteilungen [DIN Screen, Version 2. Guide line for monitoring and optimizing 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* (3rd ed.). 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 utilizados en España [A model 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 Personality 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 Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 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. Hambleton, 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.