SHL.

OPQ32r™
Technical Manual

Table of Contents

List of Tables

List of Figures

OPQ32r™ Technical Manual Chapter Summaries

This OPQ32r™ Technical Manual is intended to be read in conjunction with the OPQ32r User Manual and the OPQ Technical Manual (SHL, 2006) which covers the design, development and technical characteristics of the OPQ32i and OPQ32n. The content of the user manual focuses on administration, scoring, norming and interpretation issues, and is intended to cover all the matters one needs to refer to when using the OPQ32r. The OPQ32r Technical Manual is intended for reference purposes and provides all the technical information needed when evaluating the OPQ32r in terms of its suitability for use.

The following summaries provide outlines of the contents of each chapter.

Chapter 1: Development of OPQ32r

This chapter describes the development of the OPQ32r. First the rationale for creating the OPQ32r is outlined by discussing advantages of the forced-choice format and limitations of Classical Test Theory (CTT) for scoring forced-choice instruments. It is concluded that the problematic psychometric properties of ipsative data are not inherent in the forced-choice format itself, but originate from the CTT method of scoring which does not adequately describe the decision process of responding to forced-choice items.

The Item Response Theory approach (Thurstonian IRT) applied to the development and scoring of the OPQ32r is described, whereby ranked preferences are translated into pair-wise dichotomous outcomes, which are then linked to the set of traits measured by the test. In this approach, responses to all items are considered in the recovery of the latent traits through probabilistic estimation, rather than operating on a scale-by-scale basis. The scores obtained from the OPQ32r are no longer ipsative but have a normative data structure.

The OPQ32r was created by examining all scales of the OPQ32i using the Thurstonian IRT model, and removing a quarter of the items that provided the least information for trait estimation. The format chosen for the OPQ32r was blocks of three statements (triplets), instead of blocks of four (quads) as used in the OPQ32i. The triplet format makes the completion task quicker and less cognitively challenging.

Chapter 2: Reliability and consistency

This chapter describes the IRT based reliability of the OPQ32r scale scores and the approach underlying the Consistency score which is used for detecting random responding. The OPQ32r shows good levels of reliability even though the number of items is reduced compared to the OPQ32i. This is because the information respondents provide on forced-choice items is used more efficiently. Moreover, errors of measurement are computed for each particular set of theta scores, (i.e. for each individual profile). Data drawn from 30 country samples shows that IRT based reliabilities are also high across different language versions and countries, ranging from 0.68 to 0.94 for individual scales (median = 0.86 across all 32 scales).

Two studies are presented that demonstrate high test-retest reliability of the OPQ32r. In the first study (N=168), participants completed the OPQ32i and OPQ32r. Responses to the OPQ32i were rescored using the IRT normative model and correlations between the rescored OPQ32i and OPQ32r computed. In the second study (N=100), volunteers completed the OPQ32r on two occasions. Test-retest reliabilities were generally high (study 1: median=0.78, study 2: median = 0.85).

As several OPQ32r scales are typically combined for use in selection and development, reliabilities are provided for composite scores predicting the 20 competencies of the Universal Competency Framework (UCF). Composite reliabilities are very high, showing a median reliability of 0.91 across all 20 UCF competencies.

This chapter also describes the Consistency score which was enhanced with the release of the OPQ32r. It is based on precise measurement of probabilities of a response pattern. It is very effective in detecting random responding, and at the same time it does not penalise individuals with average trait scores, as was the case with OPQ32i Consistency.

Chapter 3: Scaling properties of OPQ32r scores

This chapter describes the scaling properties of OPQ32r scores and how they compare with OPQ32i and OPQ32n scores. A sample of 551 individuals who completed both the OPQ32n and OPQ32i instruments is used to examine relationships of OPQ32r with OPQ32n and OPQ32i. Responses to OPQ32i were rescored using IRT to produce normative OPQ32r-equivalent scores.

Results show that the IRT scoring methodology produces scale scores with good average profile location distributions comparable to the normative OPQ32n scores, but without the ipsative distortion between profiles as in OPQ32i. Moreover, the IRT scoring

methodology recovered similar score profile shapes to the normative OPQ32n scores, but without uniform response bias often present in rating scale scores. Therefore the OPQ32r is an IRT-scored forced-choice version of OPQ32 that measures the same constructs as its predecessors, but with superior resistance to response bias and score distortion.

Chapter 4: Construct validity

In this chapter several studies are presented that support the internal and external construct validity of the OPQ32r. Using the calibration sample (N=518) and the UK General Population norm sample (N=22,612), scale intercorrelations and results of exploratory factor analyses (EFA) confirmed the normative data structure of the OPQ32r scores. A third sample (training delegates who completed the OPQ32n and OPQ32i which was rescored to produce OPQ32r scores) was used to compare factor structures of OPQ32n and OPQ32r scores. The factor solutions were nearly identical, again demonstrating the normative data structure of the OPQ32r scores. Results of an EFA performed on a high stakes sample (OPQ32i data rescored to obtain OPQ32r scores) also show a clear and meaningful factor structure, recovering the Big Five Factor model as before but with a stronger Achievement factor.

This chapter also presents evidence that the OPQ32r pattern of scale intercorrelations is similar in other language versions, indicating that the 32 constructs are transferable to other cultures. Structural equation modelling (SEM) tested the level of equivalence when scale intercorrelation matrices of 29 individual countries/regions were each compared against those of the UK and US versions of the OPQ32r (overall N=118,324). Results indicate a high level of construct equivalence.

OPQ32r scales were compared to scales of the Motivation Questionnaire (SHL, 1992, 2002), cognitive ability tests (Verify verbal and numerical) and another measure of personality (Hogan's Development Survey, HDS) which focuses on "dark side traits" in a normal population. Relationships of the OPQ32r with other instruments show high correlations between scales with similar content and low correlations between scales that are different, thereby providing additional evidence for the construct validity of the OPQ32r.

Chapter 5: Criterion–related validity

Four concurrent validation studies are presented that show statistically significant relationships between predicted OPQ32r scales and line manager ratings of performance. In all four studies the strengths of hypothesized relationships exceed those of non-hypothesised ones.

The first study focuses on the OPQ32r as a predictor of management competence in a development setting. Data was drawn from OPQ32i responses (IRT-rescored to produce OPQ32r-equivalent scores) and 360-degree ratings of competencies. Using OPQ32r composite scores to predict specific leadership competencies, composite validities reach as high as 0.29 for manager ratings, 0.30 for colleagues and 0.27 for direct reports (uncorrected for range restriction and criterion unreliability).

Subsequent studies collected OPQ32r data and performance metrics for supply chain supervisors and salon employees in a large US-based retail organization, and for front line supervisors in a major manufacturer. Hypothesized relationships reached a maximum correlation of 0.26 for individual OPQ32 scales (uncorrected for range restriction and criterion unreliability). Combining OPQ32r scales to form composite scores, correlations with overall performance ratings reach an r of 0.29 (uncorrected for range restriction and criterion unreliability).

Chapter 6: Group Comparisons

To ensure the fairness of OPQ32r scores, group comparison analyses were conducted on several large datasets, including an international general population sample (N= 118,324). Gender differences were found on a number of scales, but the magnitude of these differences typically small (below one sten). The differences were of medium or small effect sizes and indicated that men described themselves as more Competitive, less Vigorous and less Caring compared than women. These results are in line with findings from the literature on gender differences on the Big Five Factor scores. Gender differences were also examined across 31 countries/regions. While the pattern of gender differences was generally consistent, the size of the differences varied across countries. This is a finding that has been observed in previous studies in the literature on gender differences in personality.

Correlations between age and OPQ32r scale scores were also small. Across all scales the average absolute correlation was 0.03.The largest correlation with age was found for the Trusting scale (0.11) indicating older employees rated themselves as more trusting compared to younger ones.

Scores across ethnicity groups were compared using the OPQ32r UK General Population Norm Sample (N=22,612). The magnitude of the differences is typically small (below one sten). Scores across different ethnic groups were also compared using a US sample (N = 2,473). Similar to the UK sample, the magnitude of the differences is typically small (below one sten).

Comparing OPQ32r scores of individuals who obtained a university degree (and potentially additional postgraduate degrees) to those without a degree, a number of small effect sizes were found, indicating that individuals with a degree were, for example, more Evaluative, more Forward thinking, more Achieving and more Data rational.

The largest group differences were found when comparing scores of managers and non-managers. A number of small and medium effect sizes were found, showing for example, that managers were more Controlling, more Decisive, less Affiliative, more Forward thinking and more Persuasive.

Chapter 7: Comparisons by country or region

Using the international norm sample (N=118,324), OPQ32r scale scores across 30 countries/regions were compared. Mean scores of individual countries or regions were compared against the overall sample mean.

Results show that mean scale score deviations from the overall average were small, typically falling within 0.20 of a standard deviation (less than 0.5 sten). Most differences were as expected; countries that are culturally similar have more similar OPQ32 profiles.

Chapter 8: Norms and comparison groups

This chapter provides illustrative examples of some of the norms available for the OPQ32r. When the OPQ32r was launched in 2009, as interim norm provisions, all previous OPQ32i norms were available through equating. Strong relationships between the OPQ32i and OPQ32r were used as a basis for linear equating the raw ipsative OPQ32i score and the OPQ32r theta score.

Following expanding use of the new OPQ32r, over 100 new norms were created since the launch, spanning more than 24 languages and 40 countries/regions. These new norms were based on up-to-date live data collected using the OPQ32r, and replaced the legacy OPQ32i equated norms. This chapter does not provide an exhaustive listing of norms available. Detailed information on norms can be obtained from the norm fact sheets and the technical documentation that was created for the norm updates (country and international norms: SHL, 2015b, 2015c). The fact sheets and technical documentation can be accessed by contacting an account manager.

Chapter 1: Development of OPQ32r

In this chapter, the rationale for making changes to the OPQ32i and the development of the OPQ32r are described.

The OPQ32 model is an occupational model of personality, which describes 32 dimensions or scales of people's preferred or typical style of behaving, thinking and feeling at work.

The OPQ32 was designed for international use and as part of its development all constructs and items were reviewed for application in different countries and cultures. It has been adapted into 30 languages and is particularly appropriate for use with professional and managerial groups, although the content of the OPQ32 is based on generic personality characteristics which are important to a wide variety of roles.

The OPQ32 model follows the general OPQ model of personality, which breaks personality down into three domains: Relationships with People, Thinking Styles and Feelings and Emotions. The three domains are joined by a potential fourth – the Dynamism domain – which is composed of scales such as Vigorous, Achieving and Competitive that relate to sources of energy. A detailed description on the development of the OPQ model and the ipsative (OPQ32i) and normative (OPQ32n) versions is available in the OPQ32 Technical Manual (2006).

1.1 Rationale for creating OPQ32r: Advantages of the forced-choice format

Prior to OPQ32r, there were two questionnaires using the OPQ model, namely the OPQ32n (*normative* or *rating* scales, using single-stimulus format) and OPQ32i (*ipsative* scales, using forced-choice format). Extensive technical documentation relating to these is provided in the OPQ32 Technical Manual (SHL, 2006) and related Technical Supplements.

Normative or rating scales have been favoured by the psychometric research tradition and are widely used in personality assessment. However, single-stimulus items are subject to numerous response biases such as acquiescence, leniency, extreme and central tendency responding (Van Herk, Poortinga, & Verhallen, 2004), idiosyncratic interpretation of the rating categories (Friedman & Amoo, 1999), halo/horn effects (Murphy, Jako & Anhalt, 1993) and socially desirable responding. These biases can be a serious threat to validity.

We were among the pioneers of the multidimensional forced-choice (MFC) format. Forced-choice measures ask respondents to choose between statements measuring different traits according to the extent to which the statements describe their preferences or behaviour.

The OPQ32i consists of 104 blocks of four statements measuring different traits. For each block respondents have to choose one item that is "Most like me" and one "Least like me". Here is an example of a block and example choices made by a respondent:

A	I like to do things my own way	☑ *most like me*
B	I recognise weak arguments	
C	I take care to follow procedures	☒ *least like me*
D	I like to spend time with others	

Because respondents cannot endorse all items, it was expected that acquiescence responding and halo effects would be eliminated, typically resulting in a greater differentiation of scores within a profile. Indeed, the forced-choice format has been proven effective against uniform response biases (Cheung & Chan, 2002), and against halo effects leading to greater operational validity coefficients (Bartram, 2007). It has also been shown that forcing to choose between seemingly equally desirable items may reduce social desirability responding (Christiansen, Burns & Montgomery, 2005), making forced-choice formats particularly attractive in candidate assessment contexts. It is commonly found that the MFC format substantially reduces score inflation compared to the single-stimulus (SS) format, at least at the group level of analysis (Christiansen et al., 2005; Jackson, Wroblewski & Ashton, 2000; Martin, Bowen & Hunt, 2002) and it is more resistant to distortion to its covariance structure (Brown, 2008; Brown & Maydeu-Olivares, 2010; Brown & Maydeu-Olivares, 2012).

1.2 Limitations of Classical Test Theory for scoring forced-choice instruments

Despite their clear advantages in reducing bias, forced-choice instruments have been criticised because their traditional scoring methodology results in *ipsative* data, which poses threats to construct validity and score interpretation as well as other substantial psychometric challenges (e.g. Dunlap & Cornwell, 1994; Meade, 2004; Tenopyr, 1988).

With classical scoring methodology, inverted rank-orders of items related to a scale are summed to produce a raw score on the scale. For example, in OPQ32i, an item selected as "most like me" contributes 2 points to the scale it measures, and the "least like me" item contributes 0 points. The items not selected as "most" or "least", add 1 point each to their respective scales. It is easy to see that while allowing for a great variability between the measured scales, item scores in the block always add up to the same number regardless of the choices made. In the OPQ32i, the total score for each block is 4. Therefore, the total score on the questionnaire – the sum of all the blocks – is the same for each individual. Below the psychometric properties of ipsative data are outlined and their implications for psychological assessment are discussed.

Relative nature of scores

Because everyone has the same total score on the test, it is impossible to get high (or low) raw scores on all scales in a multi-trait questionnaire. Therefore, some have argued (e.g. Closs, 1996), ipsative scores make sense for comparison of relative strength of traits *within* one individual, but they do not provide information on absolute (normative) trait standing, so comparisons *between* individuals are meaningless.

The fact frequently overlooked by such critics is that the number of measured traits can substantially influence the extent to which the absolute trait standings are distorted. It has been shown that with a large number (16 or more) of relatively independent scales, the impact of changes in any one trait on other traits is small (Bartram, 1996). With 30 or more measured scales, inter-individual comparisons can be performed meaningfully (Baron, 1996). Most importantly, the ordering of people on each trait largely corresponds to their normative ordering (Baron, 1996; Karpatschof & Elkjaer, 2000), and norming of ipsative scores is appropriate. A large study comparing results from the OPQ32i and OPQ32n showed that the ordering of respondents on scales derived from the two formats is indeed similar (the median scale correlation reported in the OPQ Technical Manual for a large study with training delegates was 0.71). Thus, selection decisions made using either version of OPQ32 would be similar.

Nevertheless, while allowing for a great variability of scale scores within each profile, the ipsative OPQ does not allow for variability of average profile locations. If the average profile location for normative profiles typically range from around $z=-0.8$ to $z=0.8$ with mean 0.00 and standard deviation 0.3, the ipsative profiles all centred on zero. Put simply, the scoring method places a constraint on the overall profile that makes it impossible to have high or low scores on all 32 scales. Again, how much of a problem this is depends on the number of scales measured by the test. Even though it is possible for respondents to have most of their true scores either all high or all low, the reality is that with many measured dimensions the empirical probability of such profiles is low. For example, for 30 independent traits the probability of having all true scores on the same side of the profile, that is all high or all low, is one in a million (Baron, 1996). For the normative version of OPQ32, only 0.07% of a large representative sample (N=2,951) had profiles with more than 80% of scores significantly above/below the mean (Baron, 1996). Because the true scores of most people do average at about zero, the forced-choice format would not distort their profiles too much. Nevertheless, the constraint placed on the average profile score remains the most serious theoretical limitation of ipsative data.

Distorted construct validity

It has been shown that in ipsative tests the average correlation between scales is a negative value, and approaches zero as the number of scales increases (Clemans, 1966). Again, how much of a problem this is depends on the number of scales in the questionnaire. With 32 scales, the average off-diagonal correlation is only -0.03, allowing for a wide range of both negative and positive correlations between scales (Bartram, 1996; Baron, 1996). However, scale correlations are depressed in the OPQ32i as compared to the OPQ32n, which makes it difficult to evaluate construct validity of the ipsative version directly. Moreover, conventional factor-analytic procedures are inappropriate with ipsative data. If attempted, the factor analysis extracts bipolar factors, which include contrasting scales from two different normative factors and are difficult to interpret (Baron, 1996).

Distorted reliability estimates

It is generally agreed that the forced-choice format distorts the internal consistency of instruments. Some authors have argued that appropriateness of other types of reliability, such as test-retest, is also doubtful due to violation of other common assumptions in forced-choice data, such as independence of error variance and interval level data (Meade, 2004). Regarding the internal consistency, Cronbach's alpha is an inappropriate statistic for the traditionally scored forced-choice format because ipsative data violates several assumptions, most notably consistency of coding as illustrated below.

Consistency of coding assumes that, for positively keyed items, a high item score should correspond to a high trait score. It is, however, clear that the top rank given to an item in a block does not necessarily correspond to a high *absolute* score on its respective trait – instead, it indicates a higher score *relative* to other traits. Let us imagine that an individual's true trait scores are ordered as follows:

…< trait A < trait B < trait C < **trait D** < trait E < trait F < trait G <…

Then in a block including four positively keyed items from the first four traits in the above sequence (trait A – trait D), the respondent is expected to rank an item measuring trait D as "most like me", because his/her standing on the trait is the highest out of the four traits. However, in a block including the last four traits (trait D – trait G) the respondent is expected to rank an item measuring trait D as "least like me". Therefore, the same trait will receive the highest number of points in one block, and the lowest in another. This responding, completely consistent with the true scores will appear to be inconsistent from the item coding perspective.

In OPQ32i, the number of all possible comparisons between scales is very large ($32 \times 31/2 = 496$). Consequently, items from any given scale are put in blocks with items from different traits and the opportunities for inconsistencies in coding, described above, are great. This will inevitably lead to depressed estimates of reliability.

Clearly, coefficient alpha is not a suitable basis of assessing reliabilities of an instrument using the forced-choice format, and such instruments built to ensure high coefficients alpha are almost certainly longer than they need to be. Indeed, while 6 to 8 items per scale are enough to reach acceptable reliability with the OPQ32n, as many as 13 items per scale were required to reach the same levels with the forced-choice OPQ32i. This has an implication on the time it takes to complete the test and on the experience of test takers. Moreover, alternative methods of estimating reliability in MFC tests are called for.

In summary, the problematic psychometric properties of ipsative data are not inherent in the forced-choice format itself, but originate from the current method of scoring. The traditional scoring methodology cannot adequately describe the decision-making process behind responding to multidimensional forced-choice items. Modelling this decision process correctly is the key to making the most of the information contained in forced-choice responses.

IRT as a basis for model forced-choice responding

While some still argue about the controversies of ipsative data, the focus of the debate has shifted during the last few years. Researchers have been looking into alternative ways of scoring the forced-choice format. Advances in IRT, specifically in multidimensional IRT, have made it possible to introduce models that deal with some specific types of multidimensional forced-choice measures (e.g., McCloy, Heggestad & Reeve, 2005; Stark, Chernyshenko & Drasgow, 2005).

An approach taken with the forced-choice version of the OPQ32 is based on the seminal ideas set forth by Thurstone (1927, 1931), and it draws on advances of Thurstonian factor models for ranking data (Maydeu-Olivares, 1999; Maydeu-Olivares & Bockenholt, 2005). Brown and Maydeu-Olivares (2011) introduced an approach whereby ranking preferences are translated into pair-wise dichotomous outcomes, which are then linked to a set of traits measured by the test. The **Thurstonian IRT model** describes the decision process of responding to any stimuli presented in ranking or paired comparison form, and is ideally suited to multi-trait forced-choice tests compiled of many ranking blocks of different sizes (i.e. pairs, triplets, quads etc.; see Brown & Maydeu-Olivares, 2011). The sections below give a short account of the Thurstonian IRT model relevant to the OPQ32r, including the coding protocol, item characteristic and item information functions. Further technical detail related to the model in general, such as derivation of all formulae, estimation precision of model and person parameters, and recommendations on optimal test design can be found in Brown and Maydeu-Olivares (2011).

1.3 Binary coding of forced-choice items

To enable the use of this model with blocks of four items, first it is necessary to recode responses into paired comparisons. This is the standard coding used in the Thurstonian modelling literature, as described in Maydeu-Olivares and Bockenholt (2005).

When rank-ordering statements, respondents perform mental paired comparisons of all available options, that is, every statement is compared with every other one. In effect, respondents are asking themselves: "Is statement **A** more, or less, true of me than statement **B**?" If one asks that question for item **A**, comparing it with items **B**, **C** and **D**, and then repeats the same for each of the remaining items in a block, then there are **six** ($6 = 4 \times (4-1)/2$) comparisons to make: {A, B}, {A, C}, {A, D}, {B, C}, {B, D} and {C, D}. For an item to qualify to be "most like me", it has to be compared with all remaining items and "win", or be preferred in, every

comparison. Similarly, to be "least like me" the item has to be compared with all remaining items and "lose", or be not preferred in, every comparison.

A	I like to do things my own way	☑ *most like me*
B	I recognise weak arguments	
C	I take care to follow procedures	☒ *least like me*
D	I like to spend time with others	

Responses given to a block of four statements are equivalently recoded into 6 directional paired comparisons. Each paired comparison {X, Y} is coded as follows: If item X is preferred to item Y, the outcome of the comparison is 1, otherwise it is 0. In the worked example above, item A was selected as "most like me" and item C as "least like me". It can be seen that item A is preferred in every comparison involving it, and item C is not preferred in every comparison. Because it is not known how the respondent would rank-order items B and D (the two items not selected as either most or least), this paired comparison outcome is unknown. Therefore, the outcomes of the six paired comparisons are as follows: {A, B}=1, {A, C}=1, {A, D}=1, {B, C}=1, {B, D} is unknown and {C, D}=0.

1.4 A two-dimensional IRT preference model for paired comparisons

Thurstone's theory attributes comparative judgement to the relative unobserved *utility* (psychological value judgement) of objects under comparison. Applied to a personality questionnaire, Thurstone's Law of Comparative Judgement can be formulated as follows: **A respondent prefers item A to item B if his/her utility for A is greater than for B at the time of comparison**. The utilities for items are continuous and assumed to be normally distributed in the population of respondents.

In personality questionnaires, the utilities of statements are caused by the strengths of the underlying personality traits. When respondents are forced to choose between such items, their standing on the two underlying traits will influence the utilities of the choice alternatives, and therefore, the outcome of the comparison. For example, when "I like to do things my own way" is preferred to "I take care to follow procedures" it is likely that the respondent's standing on the trait Independent Minded is higher than on the trait Rule Following. Moreover, the greater the difference in the underlying trait scores, the greater the likelihood of preferring one item to another.

Having coded the choices made in a block as pair-wise comparisons with outcomes {A, B}=1 (when **A** is preferred to **B**) or {A, B}=0 (when **B** is preferred to **A**), the likelihoods of these dichotomous responses can be linked to the underlying personality traits. The likelihood of observing the given outcome of a pair-wise comparison is expressed in terms of the strength of the two underlying traits that influence the choice made by the respondent. For positively keyed items, with an increase in the true score on the trait underlying item **A** and a decrease on the trait underlying item **B**, the probability of preferring item **A** to item **B** is non-decreasing and is influenced by: a) the respondent's standing on the two underlying traits; b) the discriminations of the two items on their underlying traits; and c) an intercept value governing the combination of the latent traits where the statements' utilities are equal.

Consider two statements *i* and *k*. It is assumed that statement *i* depends on latent trait θ_a only and that statement *k* depends on latent trait θ_b only. The Greek letter *theta* is used to denote the scale score in IRT models (often referred to as a "***theta*** score" in the IRT literature). Let y_l represent a binary variable denoting whether *i* is preferred to *k*: y_l equals 1 if *i* is preferred over *k*, and 0 otherwise. The Thurstonian IRT model is given by the following item response function (IIF) of the normal ogive:

$$P(y_l = 1 | \theta_a, \theta_b) = \Phi\left(\alpha_l + \beta_i \theta_a - \beta_k \theta_b\right), \qquad (1)$$

In this equation, β_i and β_k are discrimination parameters describing the strength of the relationship between each underlying factor and the binary outcome of pair-wise comparison, and α_l is the intercept. The latent traits are assumed to be normally distributed with unit variances and freely correlated.

For positively keyed items used in the OPQ32i, both discrimination parameters β_i and β_k are positive values. The intercept parameter α_l plays the role similar to the "difficulty" parameter in one-dimensional IRT, except that it is reversed (so one can think

of it as "easiness"). The intercept here reflects the relative population values of the utilities of the two statements under comparison: that is, how "easy" it is to prefer one statement to the other. For example, far fewer people prefer the item "I take care to follow procedures" (item **C**) to "I like to spend time with others" (item **D**) than vice versa. This would be reflected in a negative intercept value for the pair {C, D}. A negative intercept value means that it is less "easy", or more "difficult" to prefer item **C** to item **D than vice versa** in this comparison. A positive intercept value would indicate that it is relatively "easy" to prefer one item to another.

The item response function defines a surface presented in Figure 1 for a typical item pair, where the probability of preferring one item to another is plotted against two latent dimensions.

This method of coding forced-choice responses as pair-wise comparisons reflects the relative nature of preference decisions, and the IRT model takes into account the multidimensionality underlying the choices made between items. Consequently, this IRT methodology provides appropriate and precise measurement of the underlying traits. It has been shown that trait scores estimated with the IRT model have none of the problems associated with ipsative data (Brown & Bartram, 2008; Brown & Maydeu-Olivares, 2011; Brown & Maydeu-Olivares, 2012).

Figure 1: Item response function (IIF) for a pair-wise comparison.
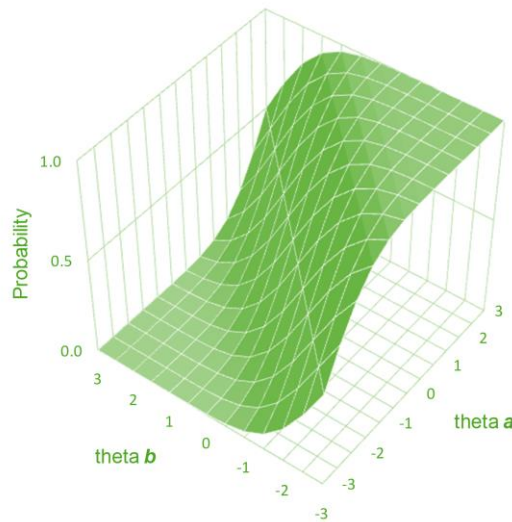
1.5 Multidimensional information and reliability

It has been found that for forced-choice questionnaires with many measured scales, reliabilities were underestimated with the coefficient alpha (Brown & Maydeu-Olivares, 2011). An advantage of using IRT is that the precision of measurement is provided by the test information function. Thus, coefficient alpha, which is an inappropriate measure when applied to traditionally scored forced-choice items, can be avoided altogether. The information function for each paired comparison is computed in a manner similar to its one-dimensional IRT counterpart, except that since each item depends on two dimensions, the direction of the information must also be considered (Reckase, 2009; Ackerman, 2005). Information on directions of the traits θ_a and θ_b measured by a pair-wise comparison {i, k} is given by the following formulae (Brown & Maydeu-Olivares, 2011):

$$\mathrm{I}_l^a\left(\theta_a,\theta_b\right)=\frac{\left[\beta_i-\beta_k\mathrm{corr}\left(\theta_a,\theta_b\right)\right]^2\left[\phi\left(\alpha_l+\beta_i\theta_a-\beta_k\theta_b\right)\right]^2}{P_l\left(\theta_a,\theta_b\right)\left[1-P_l\left(\theta_a,\theta_b\right)\right]}, \qquad (2)$$

$$\mathrm{I}_l^b\left(\theta_a,\theta_b\right)=\frac{\left[-\beta_k+\beta_i\mathrm{corr}\left(\theta_a,\theta_b\right)\right]^2\left[\phi\left(\alpha_l+\beta_i\theta_a-\beta_k\theta_b\right)\right]^2}{P_l\left(\theta_a,\theta_b\right)\left[1-P_l\left(\theta_a,\theta_b\right)\right]}, \qquad (3)$$

*where $\phi(z)$ denotes a standard normal density function evaluated at *z*. Equations (2) and (3) describe the item information functions (IIF) – in this case surfaces – an example of which is presented in Figure 2.

Figure 2: Example item information function (IIF): information in direction of trait a.
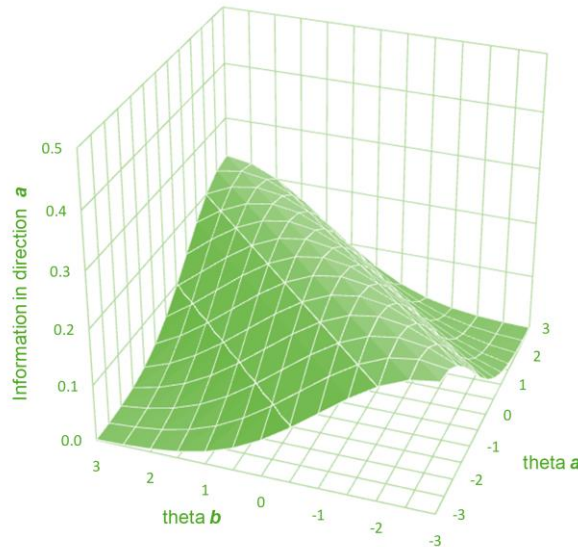
As can be seen, item information given by every pair-wise comparison is a function of the two traits measured by the two items involved in the comparison. Information also depends on the correlation between the two traits, so that the information is greater when the traits are negatively correlated than when they are positively correlated. Despite complicated dependencies of every comparison's IIFs on different pairs of traits, it is easy to compute the IIFs for a fixed set of theta scores established for an individual test taker, $\mathbf{\theta} = \left(\theta_1, \theta_2, ... \theta_p \right)$. Once the theta scores are estimated and known (how this is done will be explained below), IIFs in the direction of the first trait measured by the questionnaire can be computed for all pair-wise comparisons that involve items from that trait. It is then easy to sum all these contributions to obtain the Test Information Function (TIF) for one trait:

$$I^{a}(\mathbf{q}) = \sum_{j=1}^{N} I_{l}^{a}(\mathbf{q}). \qquad (4)$$

Because in IRT score estimation we often use prior information provided by population distribution of traits (here, multivariate normal distribution) to obtain more efficient estimates (Embretson & Reise, 2000, p. 172), this prior information is also added to the TIF to obtain the so-called posterior test information (I_p). Given a simple relationship between the information function and the error of measurement,

$$SE^{a}(\mathbf{q}) = \frac{1}{\sqrt{I_{P}^{a}(\mathbf{q})}}. \qquad (5)$$

*standard errors associated with trait *a* can be computed for the given set of individual trait scores.

Once the errors of measurement are computed for all individuals in a sample, they can be squared and then averaged to compute the average error variance for the sample. Given that we know the observed score variance in the sample (σ^2, variance of the estimated theta score), and the error variance, we can now compute *empirical* test reliability, as the proportion of variance associated with the true score (Embretson & Reise, 2000, p. 18):

$$r'_{tt} = 1 - \frac{\sigma_\theta^2}{\sigma^2} .$$ (6)

The empirical reliability provides a summary index that is useful for comparison with sample-based alpha estimates for classical normative and ipsative scores. For more details on multidimensional information, see Brown & Maydeu-Olivares (2011).

1.6 Model and person parameter estimation

To estimate item parameters, as well as trait correlations, a confirmatory factor model that incorporates relationships between the traits and the binary outcomes of comparisons, with all necessary constraints, is tested. In the OPQ32 model, binary outcomes of pair-wise comparisons serve as indicators for the 32 latent, freely correlated traits. For technical details on the identification and constraints imposed by this model, see Brown and Maydeu-Olivares (2011). The model is estimated using the general-purpose software package Mplus (Muthen & Muthen, 2003-2010).

To estimate person parameters, (i.e. trait scores for individuals), optimisation procedures are used that work not on a scale-by-scale basis, but on an item-pair-by-item-pair basis, for all scales simultaneously. Joint likelihood of observing a given response pattern is assessed for different combinations of 32 scores to find the optimum combination of scores. This is implemented through an optimisation algorithm, which maximises the joint probability of observed outcomes of comparisons (each of them is given by the respective item information function), with the added prior information given by the multivariate normal distribution with the OPQ32 covariance matrix. This is a Bayesian approach where the respondents' traits levels are estimated by maximising the mode of the posterior distributions.

1.7 Summary of the Thurstonian IRT model

To conclude,

- Responses given to questionnaires such as the OPQ can be equivalently coded as pair-wise comparisons, reflecting the relative nature of preference decisions.
- This representation enables the use of the IRT modelling that takes into account the multidimensionality underlying the choices made between items, ignored by the traditional scoring approach.
- Consequently, the IRT modelling provides appropriate and precise measurement of the underlying traits, which have none of the problems associated with ipsative data associated with the traditional scoring approach.
- Using the appropriate IRT measures of standard errors enables estimating true levels of reliability in forced-choice questionnaires, giving more accurate reliability estimates than the coefficient alpha approach.

In summary, the Thurstonian IRT model provides a means for estimation and scoring forced-choice responses, by finding the most probable combination of scale scores to explain the individual choices made in blocks of statements. "Recovery" of the latent traits underlying the responses leads to the scale scores that are no longer ipsative.

1.8 Reducing the number of items in a block

The Thurstonian item response modelling and scoring was applied to OPQ32i, and the reliability estimates derived by using the IRT information-based approach described above were around 0.9 for most scales. It is not surprising that with the more efficient use of information contained in ranking preferences, the OPQ32i with its 13 items per scale should be a very reliable instrument. Therefore, if the forced-choice version of OPQ32 was IRT-scored, the number of items could be reduced while keeping the construct breadth.

Regarding the strategy for shortening, it was possible to cut down the number of quads in the instrument, but there was another consideration. It is well known that a multidimensional forced-choice format can be cognitively challenging, particularly when more than three items are involved in one block. Processing several items at the same time requires good reading skills and comprehension, and the forced-choice format with many items in a block is generally found unsuitable for people with a low level of education. Better understanding of the decision process behind forced-choice responding offers an explanation of why it is so much more challenging to make choices in a block of four statements: this is because the number of mental comparisons to be performed is six for a block of four statements, but only three for a block of three statements. In the case of 3 items A, B and C, item A has to be compared with B and C, and item B with C, making three directional comparisons {A,B}, {A,C} and {B,C}.

If one statement is taken out of the block of four, making it a block of three, only three paired comparisons have to be performed by the respondent. This makes the completion task less cognitively challenging; therefore, the assessment potentially could be used with people with more diverse educational backgrounds. Crucially, this also offers a significant improvement in the assessment taker's experience. Of course, another added bonus is a significant reduction in the assessment completion time.

The time it takes to complete a forced-choice questionnaire is proportionate to the number of paired comparisons that need be performed. In the "most"-"least" ranking with quads, the maximum number of comparisons one performs to come to a decision is six (full ranking), and the minimum is five. Why is it five? If it so happens that the individual began with an item that then becomes the "least" (or the "most") preferred, and the second considered item was the one that would be subsequently selected as the "most" (or the "least") preferred, there will be no need to consider the pair of items that were left in the middle, thus making it only five comparisons. In many quads, however, one will have to go through comparing each item with each, even though the relative order of the two items remains unrecorded. Therefore, in the full OPQ32i with 104 quads, the number of mental comparisons one performs is between 520 (5×104) and 624 (6×104). It is easy to see that by changing to the blocks of three as compared to the blocks of four, the completion time is reduced from 3/5 up to 1/2 of the original time.

1.9 Selecting the items that provide the most information

To select the best items, each of the OPQ32 scales was carefully examined. Two samples were used to inform the selection of items for the shorter version.

Sample 1. OPQ32i Single-stimulus trial. In this trial, OPQ32i items were administered using a 5-point Likert scale. Participants volunteered and completed the questionnaires online to receive a comprehensive feedback report. Among N=632 participants, 51% were female and 49% male. The age ranged from 18 to 64 with the largest percentage of the sample coming from the 22 to 34 age group.

Sample 2. OPQ32i Standardisation sample. The OPQ32i standardisation sample consisted of 807 respondents. About two-thirds of respondents were adults working in industry and commerce, and the remaining third were students. The length of work experience was distributed as follows: 15% of respondents had work experience less than 1 year, including holiday jobs; 25% from 1 to 5 years, 25% from 6 to 15 years, and 30.4% had work experience over 15 years. Some respondents completed the questionnaire for self-development purposes, the others solely for the purposes of the standardisation study. Forty-three percent of respondents were male. Age ranged from 16 to 68, with a mean of 31 and a standard deviation of 11. For more details of this sample, see the OPQ32 Technical Manual (SHL, 2006).

First, each scale in the questionnaire was examined in relation to its dimensionality. This was done based on trials using the single-stimulus format (Sample 1). Each measured scale was examined in relation to its dimensionality and by fitting several IRT models to the data. Most scales were unidimensional, and for those, items with lowest factor loadings were highlighted for possible deletion. In several scales, a second dimension could not be ignored despite being highly correlated with the first. The second dimension typically consisted of a few items that had similar content. In those cases, items from the second dimension that did not load on the first dimension were highlighted as potential candidates for deletion. The crucial point was that after such a deletion, the scale should be no worse than it was before in measuring a coherent one-dimensional construct. In addition, the meaning of the scale should have remained the same, so it was important not to reduce the breadth of the domain measured by selecting a very narrow set of items. The common-factor model fitted to these scales after deletion of the highlighted items showed satisfactory fit.

Next, items from the forced-choice completion (Sample 2) were considered. This step is very important for two reasons. First, when put in blocks, items can interact with each other in ways that cannot be predicted from the single-stimulus presentation. Second, only actual trialling of items in blocks can establish a true picture of discrimination and intercept parameters without the confounding effects of response biases associated with rating scales. Examination of the forced-choice responses from Sample 2 carried out by fitting the Thurstonian IRT model generally confirmed the same items as in the single-stimulus trial to be problematic, and revealed a few additional items that were highlighted for deletion.

Next, a judgmental review was performed in order to remove one item from each block based on the criteria outlined above. One additional constraint was imposed: an equal number of items from each scale (3 or 4) had to be removed in order to retain 9 or 10 items per scale. This step required not only statistical information obtained from samples 1 and 2, but also detailed expert knowledge of the questionnaire's scales in order to retain items important for the construct validity of the scale. If two items highlighted for deletion were in the same block, the worst one was removed. If a block did not have any highlighted items, it was used flexibly in order to balance the number of removed items.

The final version had 104 blocks of three items (312 items), with 9 or 10 (and one scale with 11) items per scale. It is important to note that the only change made to OPQ32i was the removal of one item from each of the 104 blocks of four items. No changes were made to the remaining items nor were any changes made to the ordering of blocks of items or items within blocks. This was done to ensure that the new OPQ32r IRT scoring model could be applied retrospectively to OPQ32i item data. The equating of IRT scored OPQ32i data sets and OPQ32r is described in more detail in section 8.2.

1.10 Estimating item parameters

Sample 3. OPQ32r Calibration sample. In this trial actual blocks of three statements selected for the OPQ32r were administered. Volunteers were mainly university students from the USA, the UK and the West Indies who completed the questionnaire on paper-and-pencil (72.2%) and online (27.8%) to receive a comprehensive feedback report. Among N=518 participants, 68.9% were female and 30.3% male (0.8% did not indicate their gender). The age ranged from 18 to 55 with the largest groups from 21 to 24 (41%), 18 to 20 (21%) and 25 to 29 (14%). Ethnicity was indicated by 57% of the participants, and of these, 36% were white and 48% were Black.

Responses given to the 104 blocks of three statements were recoded into directional pair-wise comparisons as described above. To obtain estimates of the item parameters, a structural model was tested that contained 32 freely correlated latent traits (corresponding to the 32 OPQ scales), and $104 \times 3 = 312$ observed binary outcomes of pair-wise comparisons. Some additional model constraints were imposed for identification. These technical details are outside the scope of this manual and the interested reader is referred to Brown and Maydeu-Olivares (2011) for details. After the model parameters were estimated, individual scores for Sample 3 were estimated as the mode of the posterior distribution of the likelihoods of their responses.

The use of volunteers with no motivation to distort responses or engage in impression management is important at the model testing stage, particularly for models with complex structural parts. This is because the factor structure of personality-related dimensions might be distorted in high stakes contexts by additional factors (Bradley & Hauenstein, 2006; Brown, 2008), and forcing such data into the expected model will distort the item parameter estimates. Prioritising the stability of the model parameter estimates, the parameters established with this calibration sample are used to compute the theta scores for all individuals who take OPQ32r.

In summary, the shortened version of the OPQ32i was created through careful review and analysis of items and removal of the items that were providing the least information for the latent trait estimation. The shortened scales have kept the breadth of the OPQ32 traits, while becoming more coherent and unidimensional. The OPQ32r questionnaire was calibrated on a large sample. The estimated IRT parameters are used to establish the latent trait levels for individuals.

Chapter 2: Reliability and consistency

2.1 IRT-based error of measurement and empirical reliability

In Classical Test Theory, a single estimate of reliability is obtained for a scale, which is used to assess the error associated with the score. Item Response Theory offers a much more comprehensive approach to reliability through the use of test information functions. The crucial difference is that the information and consequently error of measurement in IRT actually varies depending on the scale score. As in all multidimensional IRT models, standard errors for the OPQ32r scales are computed through directional test information for particular theta values in the 32-dimensional space, and are different for every set of individual scores. However, it is often desirable to have a simple summary index that describes the level of measurement error in a particular sample. Such an index can be computed by computing the proportion of variance in the sample due to the random error, and deducing the so-called empirical reliability given by Equation (6) (outlined in section 1.5). This coefficient provides a convenient summary index for each scale and allows comparison between the reliability of the IRT-based scores and the traditional CTT scores.

Table 1 shows estimates of OPQ32r empirical reliabilities estimated from IRT information functions. The most striking result is that the IRT-based reliability estimates of the shorter OPQ32r version are still higher than alphas for the full version of OPQ32i (reported in the OPQ32 Technical manual; SHL, 2006); median increase in scale reliabilities is 0.03. Larger improvements worth noting are for the scales Outspoken, Evaluative, Forward Thinking.

It is surprising at first that the OPQ32r, with its much-reduced number of items, provides slightly higher reliabilities than the classically scored OPQ32i for most scales (median composite reliability 0.84 as compared to 0.81). The reason for it is that the IRT scoring extracts much more information from the responses by using the appropriate response model that enables optimal estimation of the latent trait scores. In addition, as explained in the section 1.2 "Limitations of the Classical Test Theory" Cronbach's alpha is an inappropriate measure of reliability of forced-choice tests.

It is important to remember that while empirical reliability provides a useful summary comparison with the classical test theory alphas, the real strength of IRT modelling is that estimates of precision are conditional on all 32 trait scores, and therefore will be different depending on individual scores on each scale. Table 1 also shows IRT-estimated standard errors of measurement for each of the 32 scales computed at the scales' mid-point (theta scores are 0 for all scales). Typically the standard error of measurement is the lowest around the middle of the distribution (where most people's scores are located), and increases as scores become more extreme. These standard error estimates are good indicators of how precise the OPQ32r really is for most people. In addition Table 1 provides results of a test-retest study in which participants completed the OPQ32r. The test-retest studies are described further below.

Table 1: Reliability estimates for the OPQ32r.

OPQ32 measured trait	Number of items	OPQ32r Calibration sample (Sample 3, N=518)		OPQ32r-OPQ32i Test-retest sample (Sample 4, N=168)	OPQ32r-OPQ32r Test-retest sample (Sample 5, N=100)
		IRT empirical reliability	Standard error (theta=0 for all scales)	Test-retest reliability	Test-retest reliability
Persuasive	10	0.83	0.36	0.77	0.85
Controlling	9	0.91	0.22	0.82	0.92
Outspoken	10	0.86	0.31	0.78	0.82
Independent Minded	9	0.77	0.41	0.73	0.86
Outgoing	9	0.89	0.25	0.84	0.92
Affiliative	10	0.84	0.33	0.82	0.80
Socially Confident	9	0.87	0.29	0.79	0.89
Modest	10	0.81	0.34	0.75	0.64
Democratic	9	0.74	0.43	0.74	0.81
Caring	10	0.81	0.37	0.75	0.82
Data Rational	10	0.88	0.26	0.83	0.88
Evaluative	9	0.80	0.39	0.69	0.79
Behavioural	10	0.79	0.39	0.78	0.79
Conventional	10	0.68	0.49	0.72	0.87
Conceptual	10	0.78	0.40	0.76	0.91
Innovative	10	0.89	0.27	0.78	0.92
Variety Seeking	9	0.77	0.40	0.69	0.81
Adaptable	10	0.87	0.28	0.82	0.74
Forward Thinking	11	0.87	0.30	0.75	0.85
Detail Conscious	10	0.89	0.24	0.78	0.86
Conscientious	10	0.84	0.35	0.80	0.78
Rule Following	10	0.89	0.26	0.79	0.79
Relaxed	10	0.87	0.28	0.79	0.82
Worrying	9	0.78	0.37	0.78	0.92
Tough Minded	9	0.80	0.39	0.69	0.88
Optimistic	10	0.81	0.37	0.78	0.91
Trusting	10	0.88	0.28	0.82	0.84
Emotionally Controlled	10	0.86	0.29	0.79	0.83
Vigorous	10	0.88	0.27	0.76	0.86
Competitive	10	0.87	0.30	0.82	0.89
Achieving	10	0.79	0.41	0.78	0.81
Decisive	10	0.83	0.35	0.73	0.87
Median		0.84	0.34	0.78	0.85

2.2 Test-retest reliability of OPQ32r with rescored OPQ32i

Some participants in OPQ32r trials also took the full OPQ32i, which provided sufficient data to compute correlations between OPQ32r scores in the two administrations. As explained in section 8.2 "Equating the OPQ32r and OPQ32i", when a subset of item responses that forms OPQ32r is scored with the IRT methodology, the resulting scores are equivalent to OPQ32r scores.

Sample 4 Volunteers completing OPQ32r and OPQ32i. Volunteers were mainly university students from the USA, the UK, and the West Indies who completed the questionnaire on paper-and-pencil (72.2%) and online (27.8%) to receive a comprehensive feedback report. The OPQ32r was administered first, followed by OPQ32i within 2 weeks from the first completion. Among N=168 participants 75% were female and 25% male. The age ranged from 18 to 55 with largest groups aged between 21 and 24 (48%), and 18 to 20 (28%). Ethnicity was indicated by only 18% of the participants, of which 88% were white.

While the OPQ32r was IRT-scored, using the parameters calibrated on the large trial sample (Sample 3, described in section 1.10), responses to the OPQ32i were used to extract the responses relevant to OPQ32r, which were then IRT-scored. Correlations between these two scores, obtained from testing the same individuals on two different occasions, provide test-retest reliability evidence for OPQ32r. These coefficients are given in Table 1 They range from 0.69 to 0.84, with the median reliability 0.78. These values are slightly smaller compared to the internal consistency values of OPQ32r reported in the same table; some of this loss can be explained by the loss of some responses to OPQ32i due to the partial ranking format used in this assessment (see section 1.3 "Binary coding of forced-choice items"). The partial ranking leads to one unknown binary outcome out of six in each block, and if both items to which this comparison relates were the ones to be retained in OPQ32r, the missing data will still be present in OPQ32r.

2.3 Test-retest reliability of OPQ32r

To examine the stability of OPQ32r scores test-retest reliability, a sample of test-retest data was collected.

Sample 5. Volunteers completing OPQ32r on two occasions. A sample of 100 Norwegian people completed the instrument in Norwegian on two occasions, with a time interval of between two days and 34 days (average 8.5 days) between administrations. The sample consisted of 23 men and 77 women. Eighty-one percent of the sample was between the ages of 21 and 29, with the remainder being between 30 and 54. Over 80% had university level qualifications. Of the 77 who reported on their employment details, all but 4 were in employment and 59 of these had been in employment for between 1 and 10 years.

Test-retest reliabilities are generally very high: median=0.85 and mean=0.84, with a maximum of 0.92. Test-retest reliabilities are above 0.80 for 25 of the 32 scales, between 0.70 and 0.80 for 6 scales, and below 0.70 for only one scale (modest, r=0.64). Scales means show very little variation between administrations.

2.4 Reliabilities of OPQ32 based composite scales

Selection decisions should not be based on single OPQ32 scales but rather on sets of OPQ scales that predict job-relevant competencies. Internal consistency reliabilities of the UCF competency factors derived from the OPQ32r scale are presented in Table 2. These reliabilities are computed using scale reliabilities and scale intercorrelations for the calibration sample. The reliabilities of the UCF composite scales have a median and average of 0.91 across all 20 competencies, ranging from 0.87 to 0.94.

Table 2: Alphas for the UCF20 competency scales derived from OPQ32r.

UCF 20 competency	Reliability
1.1 Deciding & Initiating Action	0.92
1.2 Leading & Supervising	0.91
2.1 Working with People	0.90
2.2 Adhering to Principles & Values	0.89
3.1 Relating & Networking	0.91
3.2 Persuading & Influencing	0.92
3.3 Presenting & Communicating Information	0.93
4.1 Writing & Reporting	0.91
4.2 Applying Expertise & Technology	0.91
4.3 Analysing	0.90
5.1 Learning & Researching	0.90
5.2 Creating & Innovating	0.89
5.3 Formulating Strategies & Concepts	0.88
6.1 Planning & Organising	0.93
6.2 Delivering Results & Meeting Customer Expectations	0.94
6.3 Following Instructions & Procedures	0.90
7.1 Adapting & Responding to Change	0.87
7.2 Coping with Pressure & Setbacks	0.89
8.1 Achieving Personal Work Goals & Objectives	0.91
8.2 Entrepreneurial & Commercial Thinking	0.93

Reliabilities derived for the Great Eight competencies (Bartram, 2005) and Big Five factors are presented in Table 3 and Table 4 respectively. Reliabilities are 0.86 and larger. The mapping and rationale of the Big Five factors as derived from the OPQ32 scales are described in the OPQ32 Technical Manual (2006).

Table 3: Alphas for the Great Eight competency factors derived from OPQ32r.

Great Eight competency	Reliability
Leading & Deciding	0.91
Supporting & Co-operating	0.88
Interacting & Presenting	0.92
Analysing & Interpreting	0.89
Creating & Conceptualising	0.92
Organising & Executing	0.94
Adapting & Coping	0.86
Enterprising & Performing	0.86

Table 4: Alphas for the Big Five factors derived from OPQ32r.

Big Five factor	Reliability
Extraversion	0.95
Openness	0.90
Emotional Stability	0.93
Agreeableness	0.90
Conscientiousness	0.95

2.5 Consistent responding and the Consistency score

In addition to the 32 main scales, the OPQ32r questionnaire includes a Consistency Score. The Consistency Score assesses whether an individual has responded to similar items in a consistent manner and is designed to identify random response patterns (inconsistent responding). Respondents who understand the rating task, and are motivated to answer the questionnaire honestly and accurately, tend to respond in a consistent manner when presented with similar items. Inconsistent responding, however, leads to a low Consistency Score and indicates that there may be a potential issue with the interpretation of the OPQ results.

There are a number of reasons why an individual may have responded less consistently. For example, respondents who have trouble understanding the language used, are trying to manage their impression by choosing items they think will be viewed more positively, have difficulty understanding the 'most' and 'least' rating options or are distracted when completing the questionnaire, tend to be inconsistent in their answers. The accuracy of the OPQ profile can be compromised by inconsistent responding. The purpose of the Consistency Score is to help identify such cases. If an individual receives a low Consistency Score, an OPQ trained user should explore the reasons behind this with the individual. The first consistency measure was developed with the Classical Test Theory (CTT) scored OPQ32i questionnaire. The method measures how much one's profile deviates from the 'flat' profile resulting from completely random responding. Responding completely at random means that all items in the same forced-choice question have equal chances of being selected, regardless of the scales associated. Therefore, assignment of scores to scales will be random, making a very 'flat' profile where all scales have average scores. Responding consistently, on the other hand, means that items from certain personality traits have higher (or lower) chances of being selected in line with the respondent's personality. Therefore, scale scores in the resulting profile will deviate from the average and be more differentiated. While this CTT-based Consistency Score is very effective in identifying random responding (Appendix C; OPQ32 Technical Manual; SHL, 2006), it has the tendency of favouring respondents with more differentiated profiles and penalising individuals with truly average or 'flat' profiles.

The IRT approach used to model responses to the forced-choice blocks in OPQ32r brings another perspective to understanding consistent/inconsistent responding. Under this framework, responding consistently with one's true scores means that the probability of each observed response (binary outcome of a pair-wise comparison in a triplet) should be higher than the chance level of 0.5. The probability of preferring one item to the other in a pair is determined by the respondent's standing on the two associated underlying traits. After the optimal combination of trait scores that maximises the probability of the observed response pattern for an individual is found, the probability of each observed binary outcome is evaluated to see if it is above or below 0.5. The probabilities are computed very precisely using IRT, taking into account item characteristics in the pairs and the individual's trait standings. The scoring places no importance on how likely the response is, as long as it is over the chance level of 0.5. Counting responses with a probability of above 0.5 and calculating their ratio to all responses provides the Consistency Ratio, which forms the basis of the OPQ32r Consistency Score.

The IRT-based Consistency Ratio is closely related to the old CTT-based Consistency measure, with a correlation of 0.90 (N=551 OPQ training delegates, Sample 7, which is described further below). Although they are measuring the same construct, the new IRT-based Consistency measure has a couple of additional desirable features:

- The IRT-based Consistency measure places no importance on how likely the response is, as long as it is over the chance level of 0.5. This is in contrast to the CTT-based Consistency measure, which considers not only whether there are deviations from the 'flat' profile but also how big the deviations are. The new approach therefore produces a Consistency measure with a less spread out distribution, ensuring that no undue importance is placed on the magnitude of this measure for people who just happen to have extreme true scores. With the old approach, their Consistency Scores would be much higher than that of someone who has a more 'flat' score profile. With the new approach, individuals with relatively average score profiles are more likely to obtain higher Consistency Scores.

- The IRT-based Consistency measure is based on probabilities that are computed very precisely, taking into account item properties established from large samples. In the CTT-based approach, these subtleties are not taken into account as all items are assumed to function in the same way. This precise measurement of probabilities enables more efficient use of information from a response pattern.

To sum up, the new Consistency measure is closely related but superior to the old Consistency measure. In addition to making more effective and accurate use of the response information with IRT and probabilistic methods, the new Consistency measure does not penalise respondents with 'flat' profiles.

For further technical detail and usage guidance of the Consistency Score, please refer to OPQ32r Consistency Score User Guidance (SHL, 2015a).

2.6 IRT based reliabilities for 30 country samples

Using the international norm sample, IRT-based reliabilities were computed for 30 countries and regions.

Sample 6. International norm sample: This sample was based on a total of 118,324 OPQ administrations, collected from 30 regions, covering 23 languages. Data was collected from 43 countries, including 17 countries from Europe, 7 from Asia, 5 from the Middle-East, 3 from North America 8 from South and Middle America, Australia, New Zealand and South Africa. The norm sample consisted of 61% men and 39% women. Gender distribution was similar across regions except for a small number or regions (Middle East, Korea and India) where a larger proportion of the sample was male, which was representative of the candidate pool. The modal and mean age band was 35-39. Data was used from a range of industry sectors and job levels. On regional level, industry sectors were fairly balanced for most but not all countries and regions. The availability of information on ethnicity varied greatly and was not comparable across countries and regions and therefore is not included. More information on the distributions of gender, age, education, job level, managerial role, industry and countries of the overall sample can be found in Appendix A and in the technical documentation for the international norm (SHL, 2015b).

In Table 5 summary statistics (median, mean, minimum and maximum) of reliabilities are reported for 30 countries and regions. Reliabilities for each of the OPQ scales can be found in Appendix A.

Table 5: IRT composite reliabilities for 30 country and regional samples.

Region	N	Median	Mean	Min	Max
Australia	9,120	0.85	0.84	0.68	0.93
Belgium (Dutch)	2,385	0.85	0.84	0.68	0.92
Belgium (French)	2,529	0.86	0.84	0.69	0.93
Brazil	1,006	0.87	0.85	0.70	0.93
Canada	703	0.86	0.84	0.68	0.93
China	3,322	0.86	0.85	0.69	0.93
Denmark	6,809	0.86	0.84	0.68	0.92
Finland	5,381	0.86	0.84	0.68	0.92
France	4,225	0.86	0.85	0.69	0.93
Germany, Austria and Switzerland	2,336	0.87	0.85	0.68	0.93
Greater China	564	0.86	0.85	0.68	0.93
Iceland	729	0.85	0.84	0.69	0.93
India	1,098	0.86	0.84	0.68	0.93
Italy	4,550	0.86	0.85	0.68	0.93
Korea	846	0.86	0.85	0.70	0.94
Malaysia	2,672	0.86	0.85	0.70	0.94
Middle East	1,494	0.86	0.85	0.71	0.94
Netherlands	6,227	0.85	0.84	0.68	0.93
New Zealand	2,818	0.85	0.84	0.68	0.93
Norway	4,659	0.86	0.84	0.69	0.92
Pan America	971	0.86	0.84	0.69	0.92
Poland	927	0.86	0.84	0.69	0.93
Portugal	2,248	0.86	0.84	0.70	0.93
Singapore	4,053	0.85	0.84	0.68	0.93
South Africa	4,880	0.86	0.85	0.69	0.93
Spain	695	0.86	0.85	0.70	0.94
Sweden	13,244	0.86	0.84	0.68	0.93
Turkey	1,107	0.85	0.84	0.68	0.92
UK	22,612	0.86	0.84	0.68	0.93
US	4,114	0.86	0.84	0.69	0.92

In summary, the much-reduced number of items of the OPQ32r still ensures good levels of reliability. This is because the information respondents provide on forced-choice items is used more efficiently. In addition, errors of measurement are computed for each particular set of theta scores, (i.e. for each individual profile). IRT-based reliabilities are also high across different language versions and countries, as shown for 30 country samples, ranging from 0.68 to 0.94 for individual scales, with a mean of 0.84 across all 30 countries and 32 scales (median=0.86). The new Consistency Score is very effective in detecting random responding, at the same time it does not penalise individuals with average trait scores.

Chapter 3: Scaling properties of OPQ32r scores

This section describes the scaling properties of OPQ32r and comparisons with OPQ32i and OPQ32n data. Practical guidance on interpreting scores is provided in the OPQ32r user manual, which can be obtained by contacting an account manager.

3.1 Scaling properties of OPQ32r

To evaluate how IRT-scoring changed the scaling properties of the forced-choice OPQ32 (OPQ32i), a sample of respondents who took both the OPQ32n (single-stimulus format) and OPQ32i (forced-choice format) was considered. This study was reported in the OPQ32 Technical Manual (SHL, 2006; see chapter 6, "Study with OPQ training course delegates"), where the ipsative scoring for the full OPQ32i was used to look at the relationships with the OPQ32n normative scales. In this update, the responses to OPQ32i were rescored using IRT to produce normative scores, which are equivalent to those obtained from the OPQ32r (see section 8.2 for more details). In what follows, we will refer to IRT-rescored OPQ32i as OPQ32r.

Sample 7. Training delegates sample (OPQ32n and OPQ32i).

This sample consisted of 551 individuals who participated in OPQ training courses and completed both the OPQ32n and OPQ32i instruments within a few days interval. The participants were primarily Human Resources professionals, consultants or people working in related fields. 21.3% were male, 75.4% female and 3.3% did not provide gender data.

As explained in Section 1.2, one of the properties of ipsative scores is that the average profile score (average of standardised scale scores) is centred on zero – thus it is impossible to have high or low scores on all scales of the assessment. Indeed, in this study the average profile score for the OPQ32i ipsative scores range from z=-0.07 to 0.06 with mean z=0.00 and a standard deviation of z=0.02. Despite good differentiation of scores *within* profiles, there is no differentiation *between* the profile locations as all scale scores average at zero. The IRT-scored forced-choice responses (OPQ32r) demonstrate a remarkably different pattern: average profile scores range from -0.59 to 0.57 with mean 0.00 and standard deviation 0.23. Figure 3 illustrates the distribution of the average profile score for the traditionally scored ipsative OPQ32i and its IRT-scored shortened version (OPQ32r).

Figure 3: Distribution of average profile scores in the ipsative OPQ32i and IRT-scored OPQ32r.
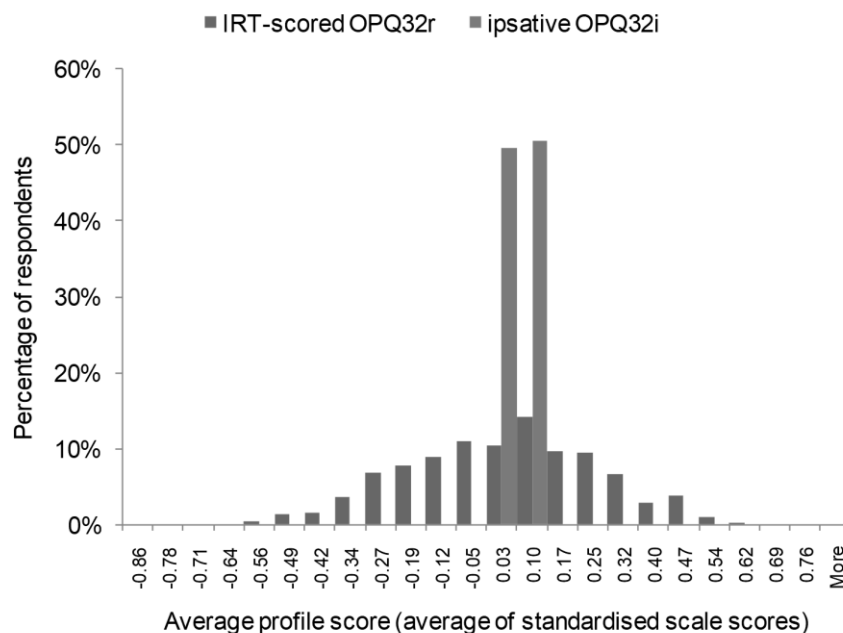
For comparison, the normative assessments' average profile scores range from z=-0.86 to 0.84 with mean 0.00 and standard deviation 0.29. Clearly, the IRT-scored forced-choice responses exhibited properties of normative scores. Figure 4 illustrates this finding by plotting the distributions of average profile scores for the OPQ32n and the IRT-scored OPQ32r.

Figure 4: Distribution of average profile scores for the normative OPQ32n and IRT-scored OPQ32r.
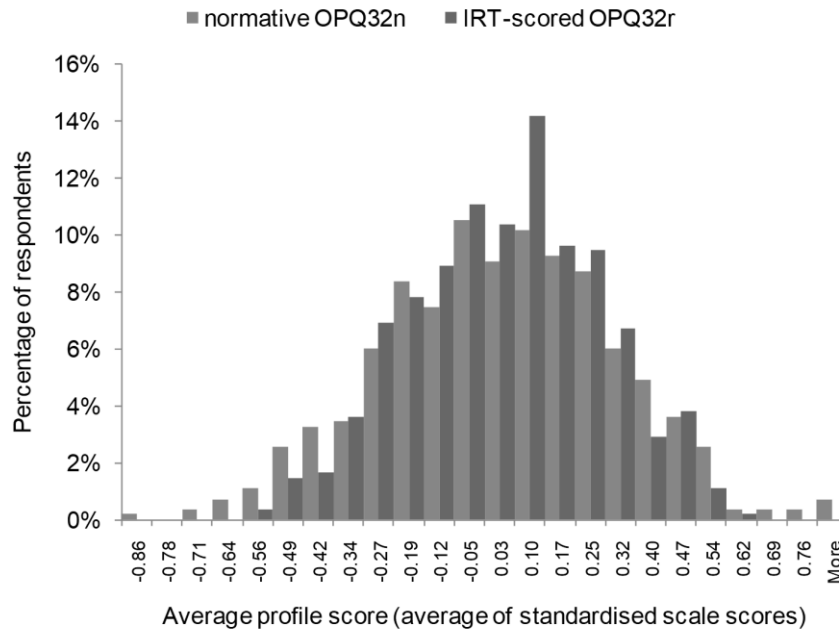
The average profile scores for the normative OPQ32n and IRT-scored OPQ32r correlate at 0.57, demonstrating that profiles "shifted" away from the mean in the same direction based on the single-stimulus and forced-choice responses.

3.2 Comparison of normative, ipsative and IRT-scored forced-choice profiles

Next, individual 32-scale profiles based on the OPQ32n normative scores and the IRT-scored OPQ32r are considered. In particular, the profiles' shape and absolute position are examined.

Similarity of profile shapes was measured by correlating 32 scale scores (normative and IRT-scored forced-choice, k=32) for the same individual. The profile similarity coefficients were distributed as shown in Figure 5. Most people (62%) had profiles with similarity 0.7 or higher, which means that self-referenced relative ordering of scales was similar based on single-stimulus and forced-choice responses. This replicates the findings reported about the OPQ32i ipsative scores – they too provide self-referenced ordering of scales that is very similar to normative.

Figure 5: Distribution of individual profile similarity coefficients (correlations between OPQ32n scores and IRT-scored OPQ32r).
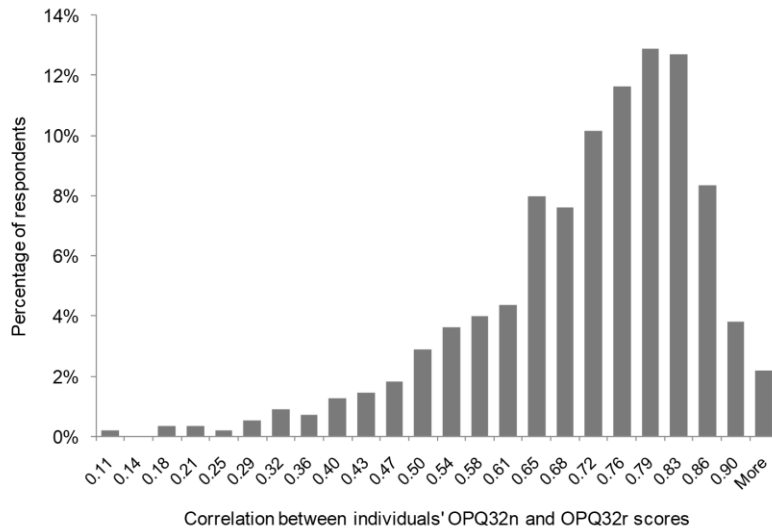
What is different about the IRT-scored forced-choice responses is that not only the shape of profiles, but also the absolute location is similar to the normative OPQ32n. The distance between profiles was measured as difference between the average of standardised normative scores and the average of forced-choice IRT scores for the 32 scales. Figure 6 shows the distribution of the profile distance scores. It can be seen that the distance or "shift" between the recovered profile and the normative profile is distributed almost normally. Most people's (98%) profiles lie within 0.5 from each other (equates to one sten), and 80% have their profiles within 0.2 (equates to 0.4 sten) or closer. Thus, absolute positions of scales are very similar based on single-stimulus and forced-choice responses.

Figure 6: Distribution of individual profile distances between OPQ32n scores and IRT-scored OPQ32r.
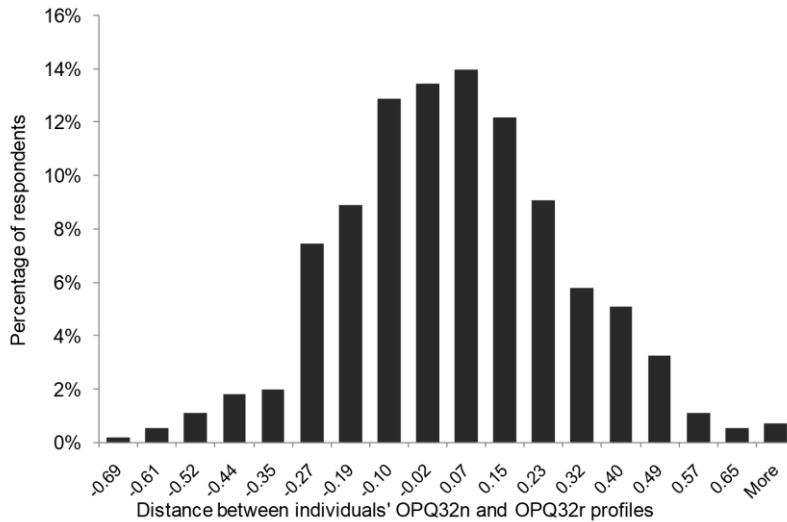
To illustrate what this means for interpretation of individual profiles, consider real OPQ32 profiles from a training course delegate. Figure 7 plots all three profiles – standardised OPQ32n normative and OPQ32i ipsative scores, and IRT-based OPQ32r scores. It can be seen that while the three profiles have similar shapes (correlation between the IRT and normative profiles is r=0.90), the IRT profile is closer to the normative (average distance d=-0.03) than the ipsative profile is (average distance d=-0.54). In fact, both the normative and the IRT forced-choice profiles are dominated by lower than average scores, having negative average profile scores of -0.48 for the normative OPQ32n and -0.45 for the IRT-scored OPQ32i. The average profile score for ipsative data, of course, can only be zero.

Figure 7: Sample individual profiles for the OPQ32n, OPQ32i and OPQ32r.
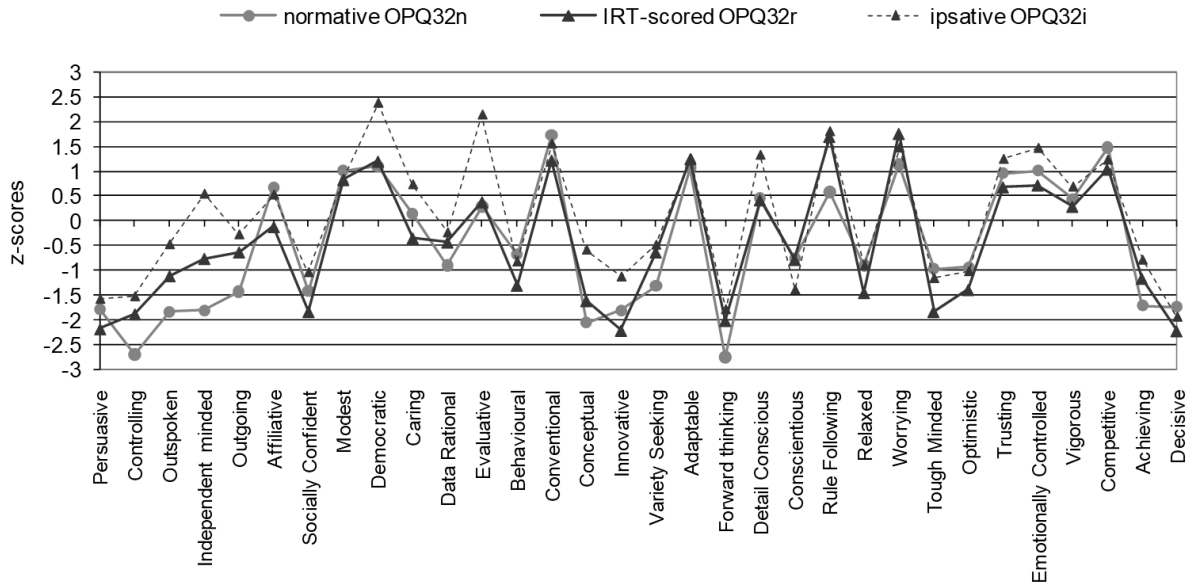
In summary, the OPQ32r scores are not ipsative. The IRT scoring allows for variance in average profile scores, and it is possible to get all high/low scores in a profile. With this constraint of ipsative data removed, comparisons between individuals can be made confidently, and the scale scores can be normed.

The IRT-based OPQ32r scores closely resemble the normative scores of OPQ32n. Importantly, the theta scores do not only resemble the shape of a normative personality profile (which the OPQ32i did well) but also its location.

Chapter 4: Construct validity

Construct validity is the extent to which a scale measures a particular hypothetical construct or trait. This is the most abstract form of validity and the most basic from a theoretical point of view. Many forms of evidence are required to build a picture of construct validity, such as intercorrelation patterns of scales within the questionnaire, factor structures and correlations with other instruments.

4.1 OPQ32r scale intercorrelations

Using OPQ32r data, it is possible to recover scale correlations of the forced-choice OPQ32 without the ipsative distortion. As explained in the OPQ32 Technical Manual (SHL, 2006), the correlation matrix for the ipsative OPQ32i is negatively biased with correlations generally lower (less positive or more negative) than their counterparts in the OPQ32n correlation matrix. This is because the ipsative scoring places a constraint on the scale variances. There is no constraint on the overall score recovered from the forced-choice ratings using IRT. Table 6 contains the full-scale correlation matrix for the OPQ32r calibration sample (Sample 3, described in section 1.10, N=518). Table 7 contains the full-scale correlation matrix for the UKE General Population norm sample (N=22,612).

4.2 Relationships with OPQ32i

Some participants in OPQ32r trials also took the full OPQ32i (Sample 4, described in section 2.2, N=168), which provided sufficient data to compute correlations between scales in the two versions. The OPQ32r was IRT-scored, using the parameters calibrated on the large trial sample (Sample 3, described in section 1.10), and the traditional ipsative scores were computed for OPQ32i.

Table 8 reports the scale correlations. The correlations between the OPQ32i and the IRT-scored OPQ32r range from 0.52 to 0.81, with median 0.71. It can be seen that these correlations are lower than the OPQ32r test-retest values based on the same sample. This is because the ipsative scale scores of OPQ32i distort the questionnaire's validity, thus reducing the relationship with its shortened IRT-scored version

Table 6: Intercorrelations of OPQ32r scales – Calibration sample (Sample 3, N = 518)

	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8	TS9	TS10	TS11	TS12	FE1	FE2	FE3	FE4	FE5	FE6	FE7	FE8	FE9	FE10
RP1 Persuasive	0.52	0.35	0.23	0.38	0.11	0.45	-0.31	0.01	0.00	0.19	0.36	0.15	-0.41	0.34	0.56	0.22	0.10	0.22	0.10	0.10	-0.29	0.16	-0.44	0.21	0.15	0.02	-0.21	0.18	0.36	0.48	0.34
RP2 Controlling	1.00	0.42	0.33	0.40	0.10	0.47	-0.30	-0.05	-0.03	0.14	0.42	0.15	-0.33	0.22	0.45	0.26	0.10	0.34	0.27	0.28	-0.20	0.06	-0.40	0.22	0.18	0.09	-0.27	0.42	0.34	0.61	0.37
RP3 Outspoken		1.00	0.40	0.47	0.13	0.39	-0.38	-0.10	-0.07	0.04	0.34	0.06	-0.23	0.20	0.30	0.21	-0.01	0.10	0.10	0.08	-0.25	0.17	-0.37	0.17	0.07	-0.04	-0.44	0.11	0.20	0.23	0.31
RP4 Independent Minded			1.00	0.15	-0.13	0.15	-0.27	-0.40	-0.15	-0.01	0.18	0.04	-0.37	0.11	0.34	0.26	0.03	0.13	0.01	0.02	-0.39	0.06	-0.24	0.04	0.01	-0.20	-0.10	0.08	0.25	0.22	0.33
RP5 Outgoing				1.00	0.47	0.60	-0.47	0.07	0.09	-0.09	0.03	0.07	-0.23	0.07	0.20	0.19	0.09	0.07	-0.06	-0.03	-0.25	0.05	-0.28	-0.02	0.18	0.12	-0.48	0.16	0.17	0.20	0.26
RP6 Affiliative					1.00	0.30	-0.26	0.32	0.34	-0.11	-0.11	0.11	-0.08	-0.03	0.00	0.06	0.01	-0.03	-0.03	0.05	-0.06	0.10	-0.06	-0.09	0.16	0.33	-0.35	0.03	-0.11	0.03	-0.01
RP7 Socially Confident						1.00	-0.28	0.04	0.19	0.15	0.26	0.12	-0.24	0.18	0.34	0.11	-0.09	0.20	0.15	0.14	-0.17	0.32	-0.60	0.27	0.31	0.26	-0.34	0.26	0.10	0.31	0.26
RP8 Modest							1.00	-0.06	-0.01	-0.09	-0.14	-0.04	0.11	-0.12	-0.24	-0.10	-0.18	-0.10	0.00	-0.04	0.16	-0.05	0.13	0.10	-0.11	-0.07	0.42	-0.13	-0.31	-0.27	-0.34
RP9 Democratic								1.00	0.36	0.00	0.07	0.28	0.03	0.16	-0.06	-0.01	0.04	0.08	0.06	0.08	0.16	-0.07	0.03	-0.12	0.05	0.30	-0.22	0.05	-0.25	0.00	-0.36
RP10 Caring									1.00	-0.16	-0.01	0.43	0.05	0.08	-0.02	-0.10	-0.13	0.14	0.13	0.18	0.21	0.00	-0.02	-0.08	0.23	0.30	-0.15	0.17	-0.38	0.03	-0.23
TS1 Data Rational										1.00	0.32	-0.08	-0.17	0.23	0.20	-0.02	0.00	0.13	0.21	0.13	-0.01	0.11	-0.20	0.13	-0.04	0.06	0.05	0.16	0.17	0.30	0.17
TS2 Evaluative											1.00	0.37	-0.25	0.54	0.47	0.17	0.06	0.32	0.37	0.28	-0.16	0.05	-0.37	0.18	0.04	-0.07	-0.15	0.22	0.10	0.45	0.23
TS3 Behavioural												1.00	-0.17	0.46	0.15	0.10	0.15	0.13	0.01	0.03	-0.12	-0.09	-0.10	-0.06	0.05	0.10	-0.13	0.07	-0.07	0.13	-0.09
TS4 Conventional													1.00	-0.37	-0.56	-0.56	-0.20	-0.14	0.10	0.12	0.53	-0.05	0.20	-0.14	-0.07	0.01	0.00	-0.14	-0.25	-0.34	-0.33
TS5 Conceptual														1.00	0.48	0.22	0.07	0.21	0.03	0.06	-0.24	0.02	-0.31	0.10	0.03	0.04	-0.17	0.14	0.05	0.34	0.12
TS6 Innovative															1.00	0.36	0.03	0.30	0.09	0.08	-0.35	0.15	-0.44	0.26	0.20	0.02	-0.17	0.25	0.21	0.51	0.36
TS7 Variety Seeking																1.00	0.14	0.09	-0.10	-0.07	-0.41	-0.04	-0.21	0.05	0.08	-0.11	-0.07	0.13	0.19	0.24	0.13
TS8 Adaptable																	1.00	-0.17	-0.09	-0.13	-0.22	-0.21	0.14	-0.15	-0.21	-0.05	-0.02	-0.02	0.14	-0.01	0.07
TS9 Forward Thinking																		1.00	0.34	0.37	0.03	-0.02	-0.24	0.06	0.36	0.02	-0.19	0.36	0.08	0.59	0.06
TS10 Detail Conscious																			1.00	0.63	0.27	0.04	-0.13	0.10	0.10	0.05	-0.10	0.41	-0.10	0.36	0.09
TS11 Conscientious																				1.00	0.30	0.00	-0.17	0.08	0.16	0.06	-0.16	0.44	0.00	0.44	0.01
TS12 Rule Following																					1.00	-0.08	0.16	-0.07	0.03	0.12	0.02	0.07	-0.27	-0.12	-0.33
FE1 Relaxed																						1.00	-0.49	0.45	0.37	0.23	-0.02	-0.09	-0.11	-0.03	0.26
FE2 Worrying																							1.00	-0.48	-0.36	-0.17	0.10	-0.19	-0.12	-0.34	-0.37
FE3 Tough Minded																								1.00	0.23	0.15	0.11	0.08	0.02	0.15	0.18
FE4 Optimistic																									1.00	0.37	-0.22	0.17	-0.05	0.25	0.09
FE5 Trusting																										1.00	-0.21	0.10	-0.18	0.02	0.00
FE6 Emotionally Controlled																											1.00	-0.14	-0.11	-0.24	-0.09
FE7 Vigorous																												1.00	0.12	0.53	0.12
FE8 Competitive																													1.00	0.39	0.27
FE9 Achieving																														1.00	0.28
FE10 Decisive																															1.00

Table 7: Intercorrelations of OPQ32r scales – UKE General Population norm sample (N = 22,612)

	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8	TS9	TS10	TS11	TS12	FE1	FE2	FE3	FE4	FE5	FE6	FE7	FE8	FE9	FE10
RP1 Persuasive	0.49	0.25	0.09	0.35	0.09	0.48	-0.33	0.10	0.08	0.07	0.24	0.24	-0.31	0.20	0.47	0.09	0.06	0.27	0.02	0.07	-0.10	0.06	-0.44	0.14	0.21	-0.03	-0.21	0.04	0.37	0.50	0.22
RP2 Controlling	1.00	0.36	0.23	0.26	0.04	0.33	-0.30	0.09	0.05	0.09	0.33	0.20	-0.31	0.13	0.41	0.20	0.02	0.35	0.04	0.09	-0.16	0.05	-0.40	0.10	0.24	0.04	-0.17	0.07	0.35	0.53	0.40
RP3 Outspoken		1.00	0.32	0.34	0.05	0.22	-0.35	0.06	-0.07	0.03	0.33	0.11	-0.23	0.21	0.25	0.19	-0.05	0.07	-0.07	-0.04	-0.21	0.05	-0.30	0.13	0.01	-0.04	-0.36	-0.05	0.16	0.18	0.27
RP4 Independent Minded			1.00	0.09	-0.11	-0.03	-0.14	-0.31	-0.18	-0.02	0.16	0.06	-0.38	0.14	0.22	0.34	0.01	0.05	-0.17	-0.17	-0.40	-0.06	-0.14	-0.04	-0.01	-0.17	-0.02	-0.05	0.19	0.12	0.26
RP5 Outgoing				1.00	0.46	0.52	-0.36	0.05	0.18	-0.20	-0.07	0.18	-0.24	0.09	0.18	0.17	0.06	-0.01	-0.14	-0.09	-0.17	0.01	-0.25	0.05	0.19	0.04	-0.36	0.07	0.14	0.19	0.12
RP6 Affiliative					1.00	0.26	-0.18	0.20	0.39	-0.18	-0.15	0.22	-0.13	-0.01	0.02	0.11	0.00	-0.04	-0.05	0.02	-0.09	0.03	-0.04	-0.08	0.16	0.28	-0.27	0.09	-0.04	0.04	-0.06
RP7 Socially Confident						1.00	-0.26	0.08	0.20	-0.02	0.09	0.19	-0.13	0.11	0.27	0.02	-0.04	0.20	0.12	0.15	0.00	0.27	-0.60	0.27	0.33	0.12	-0.23	0.11	0.14	0.31	0.11
RP8 Modest							1.00	-0.06	-0.01	-0.10	-0.16	-0.10	0.05	-0.18	-0.27	-0.02	-0.08	-0.19	-0.04	-0.06	0.00	-0.05	0.21	0.03	-0.21	0.02	0.44	-0.06	-0.30	-0.38	-0.18
RP9 Democratic								1.00	0.32	-0.03	0.18	0.33	-0.11	0.17	0.13	0.05	0.02	0.19	0.04	0.06	0.00	-0.03	-0.04	-0.04	0.10	0.26	-0.25	-0.06	-0.22	0.04	-0.19
RP10 Caring									1.00	-0.17	-0.06	0.46	-0.07	0.06	0.08	-0.03	-0.01	0.04	0.11	0.14	0.11	0.02	0.01	-0.07	0.19	0.35	-0.20	0.13	-0.28	0.02	-0.09
TS1 Data Rational										1.00	0.37	-0.13	0.02	0.11	0.11	-0.17	-0.10	0.19	0.23	0.19	0.15	0.10	-0.17	0.08	-0.01	-0.06	0.02	0.09	0.15	0.27	0.04
TS2 Evaluative											1.00	0.24	-0.25	0.42	0.39	0.11	-0.01	0.32	0.17	0.19	-0.08	0.01	-0.25	0.10	0.01	-0.05	-0.14	0.00	0.09	0.36	0.20
TS3 Behavioural												1.00	-0.34	0.37	0.21	0.17	0.18	0.17	-0.04	-0.06	-0.20	-0.08	-0.06	-0.06	0.15	0.14	-0.20	-0.02	-0.09	0.12	-0.01
TS4 Conventional													1.00	-0.38	-0.46	-0.61	-0.17	-0.19	0.30	0.22	0.54	0.08	0.09	-0.08	-0.15	-0.04	0.03	0.02	-0.15	-0.29	-0.30
TS5 Conceptual														1.00	0.45	0.17	0.04	0.18	-0.07	-0.08	-0.18	-0.01	-0.21	0.06	0.03	-0.04	-0.18	-0.08	-0.04	0.21	0.04
TS6 Innovative															1.00	0.25	-0.01	0.34	-0.07	0.02	-0.19	0.02	-0.37	0.09	0.22	0.02	-0.21	0.02	0.12	0.48	0.29
TS7 Variety Seeking																1.00	0.11	0.04	-0.35	-0.27	-0.48	-0.10	-0.06	-0.01	0.05	-0.01	-0.01	-0.03	0.07	0.10	0.19
TS8 Adaptable																	1.00	-0.10	-0.09	-0.17	-0.16	-0.17	0.13	-0.13	-0.08	-0.10	0.01	-0.07	0.01	-0.02	-0.03
TS9 Forward Thinking																		1.00	0.21	0.20	0.00	0.03	-0.29	0.04	0.38	0.06	-0.15	0.09	0.14	0.52	0.12
TS10 Detail Conscious																			1.00	0.66	0.46	0.15	-0.10	0.04	0.08	-0.03	-0.01	0.35	-0.07	0.19	-0.10
TS11 Conscientious																				1.00	0.41	0.12	-0.14	0.07	0.10	0.03	-0.05	0.41	-0.02	0.29	-0.06
TS12 Rule Following																					1.00	0.09	-0.02	-0.01	-0.04	-0.01	0.01	0.18	-0.12	-0.02	-0.26
FE1 Relaxed																						1.00	-0.42	0.38	0.31	0.15	0.04	0.04	-0.02	0.00	0.07
FE2 Worrying																							1.00	-0.38	-0.32	-0.05	0.07	-0.06	-0.27	-0.39	-0.29
FE3 Tough Minded																								1.00	0.12	0.07	0.14	0.04	0.00	0.08	0.05
FE4 Optimistic																									1.00	0.31	-0.16	0.10	0.06	0.30	0.10
FE5 Trusting																										1.00	-0.12	0.00	-0.20	-0.04	0.02
FE6 Emotionally Controlled																											1.00	0.01	-0.05	-0.22	-0.07
FE7 Vigorous																												1.00	-0.03	0.30	-0.01
FE8 Competitive																													1.00	0.47	0.15
FE9 Achieving																														1.00	0.20
FE10 Decisive																															1.00

Table 8: OPQ32r correlations with ipsative OPQ32i and normative OPQ32n.

	OPQ32r with OPQ32i (Sample 4, N=168)	OPQ32r with OPQ32n (Sample 7, N=551)
Persuasive	0.69	0.76
Controlling	0.78	0.74
Outspoken	0.73	0.69
Independent Minded	0.56	0.56
Outgoing	0.81	0.80
Affiliative	0.80	0.70
Socially Confident	0.76	0.75
Modest	0.76	0.68
Democratic	0.68	0.58
Caring	0.69	0.60
Data Rational	0.80	0.80
Evaluative	0.56	0.63
Behavioural	0.78	0.62
Conventional	0.52	0.74
Conceptual	0.63	0.74
Innovative	0.70	0.80
Variety Seeking	0.61	0.66
Adaptable	0.79	0.65
Forward Thinking	0.63	0.67
Detail Conscious	0.63	0.77
Conscientious	0.79	0.67
Rule Following	0.76	0.70
Relaxed	0.76	0.70
Worrying	0.75	0.73
Tough Minded	0.61	0.71
Optimistic	0.69	0.73
Trusting	0.80	0.66
Emotionally Controlled	0.76	0.76
Vigorous	0.70	0.61
Competitive	0.76	0.71
Achieving	0.69	0.74
Decisive	0.67	0.72
Median	*0.71*	*0.70*

4.3 Relationships with OPQ32n

Here again the sample of OPQ training course delegates is considered (Sample 7, described in section 3.1, N=551), who took both the OPQ32n (single-stimulus format) and the OPQ32i (forced-choice format). To evaluate what relationships there would be between the OPQ32n and the IRT-scored OPQ32r, the OPQ32i item responses were recoded as paired comparisons, and the IRT scoring was applied to the reduced number of items (the same items that formed the OPQ32r).

The ordering of respondents based on the single-stimulus and forced-choice responses was similar. Correlations between the normative and the IRT-scored OPQ32r ranged from 0.56 to 0.80 with median 0.70 (see Table 8). These relationships are

remarkably similar to the OPQ32n-OPQ32i relationships reported in the OPQ32 Technical Manual (SHL, 2006; chapter 7 on Construct Validity). Evidently, construct relationships with its sibling measure, OPQ32n, is preserved in the IRT-scored OPQ32r.

4.4 Exploratory factor analyses of OPQ32r

Establishing the factor structure is important to construct validity in order to show whether the relationships between scales concur with what would be expected of the constructs they are intended to measure. It has been shown in the OPQ32 Technical Manual (SHL, 2006) that the Big Five factors of personality are found in all OPQ32 normative samples. The Big Five factors are typically embedded in a larger factor solution because the OPQ32 also contains scales that do not relate conceptually to any of the Five Factor Model (FFM) factors. The OPQ32 measures a broader personality domain than the FFM. For example, energy, drive and interests elements are included in the OPQ32, but are not apparent within the Big Five.

Four independent exploratory factor analyses (EFA) of the IRT-scored forced-choice OPQ32r scales will be reviewed in this section. It is important to note that the ipsative OPQ32i presents a problem for factor analysis, typically producing bipolar factors, which contrast scales from two different normative factors. These factors reflect the traditional scoring of the forced-choice format, where receiving points on one scale inevitably means losing points on another. In contrast, conventional factor-analytical procedures can be applied to the IRT scale scores of the OPQ32r, because they are no longer ipsative.

To make the analyses comparable, a criterion for the number of factors extracted was set to be Cattell's scree test (Cattell, 1966). Generally, a large drop in eigenvalues was observed after six or seven factors. For comparison, five or six factors are typically extracted from the normative OPQ32n, five of which represent the Big Five descriptions. The sixth dimension, if extracted, is not consistent across samples (SHL, 2006). For consistency with results reported in the OPQ32 Technical Manual (SHL, 2006) in Part 2 of Chapter 7, principal component extraction with Varimax rotation was used here. The orthogonal Varimax rotation minimises the number of variables that have high loadings on each factor and simplifies the interpretation of the factors. Below, principal component analyses results for samples described previously are summarised.

Factor structure of the OPQ32r Calibration sample (Sample 3, described in section 1.10, N=518). Principal component analysis was performed on the IRT scores estimated from the forced-choice ratings. The scree plot suggested existence of six components (explaining 58.1% of the variance). The solution (see Table 9) resembles the FFM model but separates two facets of Openness: openness to new ideas or *unconventionality*, comprising of the traits Variety Seeking, Conventional (reversed) and Rule Following (reversed); and *critical thinking* (comprising of traits Conceptual, Evaluative and Data Rational).

Table 9: Rotated component matrix for the OPQ32r calibration sample (Sample 3, N=518).

	Openness (unconventionality)	Extraversion	Conscien-tiousness	Emotional Stability	Agreeable-ness	Openness (critical thinking)
Persuasive	*0.39*	**0.41**	0.19	0.20	-0.07	0.35
Controlling	*0.33*	*0.44*	**0.47**	0.13	-0.15	0.21
Outspoken	0.22	**0.58**	0.09	0.17	-0.20	0.16
Independent Minded	**0.49**	0.18	0.14	0.04	*-0.43*	-0.07
Outgoing	0.20	**0.81**	0.00	0.08	0.06	-0.13
Affiliative	-0.02	**0.57**	-0.08	0.07	*0.44*	-0.22
Socially Confident	0.15	**0.57**	0.18	*0.48*	0.11	0.11
Modest	-0.07	**-0.73**	-0.06	0.10	0.11	-0.12
Democratic	-0.12	0.13	0.01	-0.12	**0.73**	0.15
Caring	-0.02	0.06	0.21	0.07	**0.72**	-0.18
Data Rational	-0.11	0.00	0.14	0.13	-0.16	**0.64**
Evaluative	0.27	0.06	*0.36*	0.07	0.07	**0.68**
Behavioural	*0.36*	0.02	0.06	-0.14	**0.59**	0.27
Conventional	**-0.79**	-0.10	-0.01	-0.07	0.03	-0.15
Conceptual	**0.45**	0.02	0.10	0.03	0.30	**0.59**
Innovative	**0.61**	0.16	0.25	0.25	-0.04	*0.33*
Variety Seeking	**0.72**	0.06	0.06	-0.04	-0.03	-0.08
Adaptable	0.18	0.18	-0.22	**-0.43**	-0.01	0.27
Forward Thinking	0.20	0.01	**0.71**	0.09	0.09	-0.01
Detail Conscious	-0.28	0.00	**0.69**	0.08	0.01	0.23
Conscientious	-0.24	0.04	**0.76**	0.05	0.05	0.10
Rule Following	**-0.69**	-0.18	0.27	-0.02	0.18	-0.07
Relaxed	-0.04	0.09	-0.15	**0.80**	-0.03	0.07
Worrying	-0.25	-0.23	-0.18	**-0.69**	0.04	-0.27
Tough Minded	0.09	-0.11	0.04	**0.70**	-0.12	0.20
Optimistic	0.13	0.13	*0.31*	**0.55**	0.22	-0.30
Trusting	-0.16	0.23	-0.01	*0.39*	**0.46**	-0.03
Emotionally Controlled	0.02	**-0.70**	-0.18	0.05	-0.20	0.02
Vigorous	0.09	0.14	**0.70**	-0.01	0.03	0.03
Competitive	0.25	0.29	0.11	-0.18	**-0.51**	0.15
Achieving	*0.33*	0.22	**0.72**	0.04	-0.09	0.24
Decisive	0.23	*0.34*	0.04	0.28	**-0.46**	0.21

Main factor loadings are in **bold**, secondary loadings above 0.3 are in ***bold italic***.

Factor structure of the OPQ32r UKE General Population Norm Sample (N=22,612). Principal component analysis was performed on the UKE General Population norm sample. The scree plot suggested six components (explaining 54.3% of the variance). The solution is highly comparable to the solution found using the calibration sample. It resembles the FFM model but separates two facets of Openness: openness to new ideas or *unconventionality*, comprising of the traits Independent Minded, Variety Seeking, Conventional (reversed), Rule Following (reversed) and Decisive; and *critical thinking* (comprising of traits Conceptual, Evaluative, Data Rational and Innovative).

Table 10: Rotated component matrix for the OPQ32r UKE general population norm sample (N=22,612).

	Openness (unconventionality)	Extraversion	Agreeable-ness	Openness (critical thinking)	Conscien-tiousness	Emotional Stability
Persuasive	.243	**.535**	.005	.193	***.305***	.173
Controlling	***.345***	**.425**	-.059	.242	***.382***	.155
Outspoken	.171	**.574**	-.118	.295	-.200	.125
Independent Minded	**.544**	.121	***-.345***	.100	-.006	-.048
Outgoing	.204	**.719**	.199	-.277	-.009	.075
Affiliative	.063	.383	**.513**	***-.350***	.059	-.043
Socially Confident	.055	**.539**	.214	-.058	.212	***.489***
Modest	.037	**-.723**	.070	-.137	-.111	.060
Democratic	-.104	.088	**.629**	***.335***	-.063	-.071
Caring	-.056	.081	**.748**	-.082	.175	-.038
Data Rational	-.220	-.043	-.275	**.490**	.198	.122
Evaluative	.087	.106	-.033	**.794**	.132	.046
Behavioural	.266	.162	**.586**	.283	.043	-.157
Conventional	**-.807**	-.072	-.161	-.229	-.107	.032
Conceptual	.225	.151	.207	**.643**	-.132	-.001
Innovative	***.397***	.281	.096	**.483**	.225	.129
Variety Seeking	**.742**	.000	.081	.017	-.009	-.080
Adaptable	.194	.064	.039	-.033	-.042	**-.366**
Forward Thinking	.110	.090	.098	***.409***	**.504**	.112
Detail Conscious	**-.547**	-.047	.018	.163	***.537***	.076
Conscientious	***-.464***	-.004	.043	.130	**.618**	.086
Rule Following	**-.758**	-.040	-.013	-.016	.210	.061
Relaxed	-.111	.012	.042	-.026	-.009	**.753**
Worrying	-.114	***-.374***	.061	-.209	-.202	**-.677**
Tough Minded	.048	-.042	-.042	.089	-.047	**.705**
Optimistic	.157	.155	***.311***	-.065	***.364***	**.401**
Trusting	.022	-.059	**.566**	-.101	.032	.270
Emotionally Controlled	.127	**-.669**	-.252	-.163	.071	.161
Vigorous	-.103	-.032	.045	-.152	**.648**	-.022
Competitive	.198	***.404***	**-.503**	-.016	.258	-.005
Achieving	.185	***.385***	-.105	***.312***	**.670**	.047
Decisive	**.411**	.194	-.232	.089	.118	.197

Main factor loadings are in **bold**, secondary loadings above 0.3 are in ***bold italic***.

Factor structure of the Training delegates sample (Sample 7, described in section 3.1, N=551). In this low-stakes occupational sample described earlier, the respondents took both the OPQ32i and OPQ32r as a part of their OPQ training course. To score the forced-choice responses using the IRT approach, the established parameters were applied to the reduced number of items making the OPQ32r.

The scree plot suggested extracting six components explaining 58.7% of the variance. The solution clearly resembles the Five Factor Model, but just as in the solution for the OPQ32r calibration sample, it separates Openness into two facets: *unconventionality* and *critical thinking* (see Table 11).

Table 11: Rotated component matrix for the OPQ Training delegates sample (Sample 7, N=551).

	Openness (unconventionality)	Extraversion	Emotional Stability	Agreeableness	Openness (critical thinking)	Conscientiousness
Persuasive	*0.36*	**0.52**	*0.30*	-0.07	0.19	0.02
Controlling	*0.32*	**0.50**	0.28	-0.23	0.20	0.18
Outspoken	0.13	**0.71**	0.17	-0.17	0.08	-0.18
Independent Minded	*0.38*	0.30	0.05	**-0.40**	0.24	-0.22
Outgoing	0.16	**0.67**	0.17	0.15	*-0.40*	0.01
Affiliative	0.04	0.29	-0.03	**0.52**	*-0.50*	0.12
Socially Confident	0.10	*0.38*	**0.63**	0.17	-0.18	0.06
Modest	-0.05	**-0.72**	0.00	0.02	-0.11	-0.04
Democratic	-0.04	-0.01	-0.01	**0.71**	0.00	0.11
Caring	-0.08	0.08	-0.01	**0.75**	-0.18	0.02
Data Rational	-0.13	-0.05	0.08	*-0.33*	**0.56**	0.11
Evaluative	0.11	0.18	0.04	-0.08	**0.79**	0.08
Behavioural	0.24	0.07	-0.03	**0.67**	*0.37*	-0.12
Conventional	**-0.85**	-0.12	-0.12	-0.01	-0.19	0.08
Conceptual	*0.33*	0.08	-0.02	0.18	**0.67**	-0.22
Innovative	**0.61**	0.26	0.21	-0.09	*0.34*	-0.06
Variety Seeking	**0.80**	0.07	-0.02	-0.03	-0.09	0.05
Adaptable	0.20	-0.10	-0.28	0.04	-0.08	-0.03
Forward Thinking	0.33	-0.02	0.19	0.04	**0.44**	*0.38*
Detail Conscious	*-0.45*	-0.10	0.00	0.06	0.18	**0.65**
Conscientious	-0.26	0.02	0.00	0.02	0.07	**0.80**
Rule Following	**-0.69**	-0.06	-0.11	0.13	-0.04	*0.36*
Relaxed	0.01	-0.06	**0.80**	0.00	-0.01	-0.03
Worrying	-0.26	-0.23	**-0.75**	0.15	-0.11	-0.02
Tough Minded	0.05	-0.07	**0.71**	-0.15	0.18	-0.05
Optimistic	0.22	0.08	**0.65**	0.24	-0.05	0.08
Trusting	-0.14	0.00	*0.40*	**0.52**	-0.10	-0.07
Emotionally Controlled	0.09	**-0.73**	0.09	-0.24	0.07	0.02
Vigorous	0.11	-0.03	0.00	0.06	-0.19	**0.74**
Competitive	0.20	*0.42*	-0.01	**-0.51**	0.14	-0.03
Achieving	**0.50**	*0.37*	0.12	-0.17	*0.33*	*0.48*
Decisive	*0.32*	*0.30*	0.29	**-0.44**	0.01	-0.06

Main factor loadings are in **bold**, secondary loadings above 0.3 are in ***bold italic***.

To explore how similar the factor solutions are for the OPQ32n and OPQ32r, the principal component analysis was performed on the normative OPQ32n scores for this sample. As can be seen in Table 12, the solution is almost identical to the one for the IRT-scored forced-choice responses.

Table 12: Rotated component matrix for OPQ32n normative scores: OPQ Training delegates sample (Sample 7, N=551).

	Openness (unconventionality)	Extraversion	Emotional Stability	Agreeableness	Openness (critical thinking)	Conscientiousness
Persuasive	0.27	**0.47**	0.28	-0.07	*0.37*	0.15
Controlling	0.23	**0.43**	0.23	-0.17	0.22	*0.37*
Outspoken	0.24	**0.68**	0.07	-0.25	0.05	0.13
Independent Minded	**0.45**	0.19	0.05	**-0.45**	0.17	0.10
Outgoing	0.14	**0.73**	0.12	0.28	0.00	0.09
Affiliative	-0.01	0.37	0.01	**0.51**	-0.22	0.09
Socially Confident	0.12	*0.46*	0.54	0.21	0.05	0.09
Modest	-0.01	**-0.62**	-0.12	0.08	-0.19	0.00
Democratic	-0.15	0.04	0.01	**0.63**	0.18	0.11
Caring	-0.04	0.01	0.09	**0.73**	-0.04	0.19
Data Rational	-0.26	-0.09	0.20	*-0.38*	0.39	0.20
Evaluative	-0.08	0.05	0.00	-0.27	**0.65**	*0.32*
Behavioural	0.23	0.07	0.04	*0.41*	0.51	-0.02
Conventional	**-0.83**	-0.03	-0.12	0.00	-0.17	-0.03
Conceptual	0.16	0.04	0.09	-0.03	**0.69**	-0.07
Innovative	**0.48**	0.19	0.14	-0.04	*0.47*	0.04
Variety Seeking	**0.78**	0.14	0.11	0.06	0.09	0.14
Adaptable	0.06	0.02	-0.11	0.00	**0.39**	-0.10
Forward Thinking	0.11	-0.11	0.20	0.05	**0.46**	*0.37*
Detail Conscious	*-0.52*	-0.09	0.03	0.05	0.03	**0.59**
Conscientious	-0.27	-0.04	0.03	0.12	-0.04	**0.74**
Rule Following	**-0.64**	-0.11	-0.12	0.23	-0.09	0.20
Relaxed	-0.04	0.00	**0.80**	0.03	0.02	-0.04
Worrying	-0.16	-0.23	**-0.67**	0.10	-0.01	-0.02
Tough Minded	0.08	-0.18	**0.67**	-0.16	0.03	0.20
Optimistic	0.17	0.13	**0.66**	0.21	0.06	0.03
Trusting	0.02	0.04	**0.40**	*0.37*	-0.03	-0.18
Emotionally Controlled	0.00	**-0.70**	0.19	-0.11	0.17	0.15
Vigorous	0.23	0.13	-0.03	0.16	-0.12	**0.68**
Competitive	0.08	*0.33*	-0.07	**-0.52**	0.25	0.08
Achieving	0.24	0.28	0.06	-0.11	0.28	**0.59**
Decisive	*0.40*	*0.41*	0.13	**-0.42**	-0.09	-0.02

Main factor loadings are in **bold**, secondary loadings above 0.3 are in ***bold italic***.

Factor structure of the OPQ32i high-stakes online sample. It is important to explore the factor structure underlying the responses to the OPQ32i completed not only in low-stakes research settings but also in real high-stakes situations. As the OPQ32r widely used in this context, making sure that construct validity is not compromised by possible presence of impression management effects is important. The large occupational sample representative of managerial and professional candidates from a wide range of organisations across the UK was used here (Sample 8, described in section 8.2).

Principal components analysis with Varimax rotation was performed on the IRT scores estimated from the forced-choice ratings. The scree plot suggested existence of five large components and two smaller components (seven components explained 58.6% variance in total). The solution is presented in Table 13. This solution has a very clear *achievement* factor, which is separate from Conscientiousness and is indicated by the traits Achieving, Competitive, Forward Thinking, Persuasive and Controlling. As in the factor solutions discussed previously, Openness is represented by two facets: *unconventionality* and *critical thinking*. Overall, the factor structure is clear and factors present in low-stakes assessments are also found here.

Table 13. Rotated component matrix for the IRT-scored responses to the OPQ32i: High stakes online sample (Sample 8, N=18,423).

	Achievement	Openness (unconven-tionality)	Agreeable ness	Extraversi on	Emotional Stability	Openness (critical thinking)	Conscien-tiousness
Persuasive	**0.68**	0.11	0.08	0.26	0.16	0.06	-0.09
Controlling	**0.62**	0.21	-0.07	0.23	0.16	0.15	0.07
Outspoken	0.02	0.19	-0.25	**0.65**	0.23	*0.31*	-0.01
Independent Minded	-0.02	**0.55**	*-0.45*	0.19	0.01	0.17	0.06
Outgoing	0.16	0.22	0.13	**0.71**	0.10	-0.28	-0.05
Affiliative	-0.02	0.12	**0.46**	*0.45*	-0.03	*-0.33*	0.12
Socially Confident	*0.34*	0.04	0.24	*0.41*	**0.51**	-0.11	0.05
Modest	*-0.41*	0.06	0.03	**-0.61**	0.06	-0.10	0.03
Democratic	0.07	-0.11	**0.72**	0.07	-0.07	0.21	-0.04
Caring	-0.12	0.06	**0.73**	0.15	-0.01	-0.06	0.18
Data Rational	0.23	-0.23	-0.21	-0.15	0.07	**0.50**	0.14
Evaluative	0.23	0.11	0.00	0.08	0.07	**0.77**	0.13
Behavioural	0.07	0.26	**0.61**	0.15	-0.11	*0.31*	-0.06
Conventional	-0.30	**-0.77**	-0.17	-0.02	0.00	-0.19	0.06
Conceptual	0.02	0.25	0.22	0.11	-0.02	**0.67**	-0.13
Innovative	**0.43**	**0.42**	0.12	0.13	0.12	*0.38*	-0.02
Variety Seeking	0.09	**0.77**	0.04	0.00	-0.04	-0.01	-0.06
Adaptable	0.14	0.08	0.07	0.01	**-0.38**	-0.04	-0.23
Forward Thinking	**0.59**	0.05	0.19	-0.10	0.08	*0.32*	0.17
Detail Conscious	0.02	*-0.43*	0.03	-0.05	0.09	0.23	**0.66**
Conscientious	0.14	-0.30	0.03	-0.02	0.11	0.14	**0.75**
Rule Following	-0.08	**-0.69**	0.04	-0.05	0.01	0.01	*0.34*
Relaxed	0.05	-0.04	0.06	0.00	**0.77**	-0.01	0.02
Worrying	*-0.39*	-0.06	0.04	-0.20	**-0.70**	-0.11	-0.04
Tough Minded	0.05	0.03	-0.03	-0.09	**0.72**	0.09	-0.05
Optimistic	*0.40*	0.10	*0.39*	0.01	**0.41**	-0.17	0.05
Trusting	0.01	-0.05	**0.52**	-0.04	0.29	-0.11	-0.01
Emotionally Controlled	-0.13	0.06	-0.23	**-0.70**	0.12	-0.13	0.02
Vigorous	0.07	0.10	0.06	0.01	-0.02	-0.16	**0.78**
Competitive	**0.67**	0.01	*-0.41*	0.10	-0.07	-0.10	-0.15
Achieving	**0.79**	0.16	-0.01	0.11	0.01	0.19	0.26
Decisive	0.23	**0.34**	*-0.32*	0.20	0.25	0.00	0.05

Main factor loadings are in **bold**, secondary loadings above 0.3 are in ***bold italic***.

4.5 Summary of OPQ32r factor structure

There are clear commonalities in the results of the three exploratory studies presented above, despite the differences between them in terms of population, ethnic composition, age and purpose of assessment. In order to summarise and compare the results of the principal component analyses, loadings above +/-0.4 are given in Table 14. Only loadings found in at least two samples are included in the table. The components clearly represent the Big Five factors of personality, namely Extraversion, Conscientiousness, Agreeableness, Emotional Stability and Openness to Experience (McCrae and Costa, 1987), with an addition of two facets of Openness. The *achievement* factor found only in the online high-stakes sample is not included in the table.

Congruence between the data sets is clear, with the majority of scales showing consistently strong loadings on the factors. It can be seen that the scales loading on each of the FFM factors correspond well with the hypotheses developed from the expert mapping of the OPQ32 onto the FFM (described in the OPQ32 Technical Manual, chapter 7; SHL, 2006). Overall, the results support the construct validity of the OPQ32r as an instrument that may be applied to measure the FFM.

Table 14: Principal Component Analysis loadings above +/−0.4 for the three data sets.

Big Five factor	OPQ32 scales	OPQ32r Calibration sample (Sample 3, N=518)	Training delegates sample (Sample 7, N=551)	Online high-stakes sample (Sample 8, N=18,423)
Extraversion	Outgoing	0.81	0.67	0.71
	Emotionally Controlled	-0.70	-0.73	-0.70
	Outspoken	0.58	0.71	0.65
	Modest	-0.73	-0.72	-0.61
	Affiliative	0.57		0.45
	Socially Confident	0.57		0.41
	Controlling	0.44	0.50	
	Persuasive	0.41	0.52	
Agreeableness	Caring	0.72	0.75	0.73
	Democratic	0.73	0.71	0.72
	Behavioural	0.59	0.67	0.61
	Competitive	-0.51	-0.51	-0.41
	Trusting	0.46	0.52	0.52
	Decisive	-0.46	-0.44	
	Affiliative	0.44	0.52	0.46
	Independent Minded	-0.43	-0.40	-0.45
Conscientiousness	Conscientious	0.76	0.80	0.75
	Detail Conscious	0.69	0.65	0.66
	Vigorous	0.70	0.74	0.78
	Forward Thinking	0.71		
	Achieving	0.72	0.48	
Emotional Stability	Relaxed	0.80	0.80	0.77
	Worrying	-0.69	-0.75	-0.70
	Tough Minded	0.70	0.71	0.72
	Optimistic	0.55	0.65	0.41
	Socially Confident	0.48	0.63	0.51
Openness (unconventionality)	Conventional	-0.79	-0.85	-0.77
	Variety Seeking	0.72	0.8	0.77
	Rule Following	-0.69	-0.69	-0.69
	Innovative	0.61	0.61	0.42
	Independent Minded	0.49		0.55
Openness (critical thinking)	Evaluative	0.68	0.79	0.77
	Data Rational	0.64	0.56	0.50
	Conceptual	0.59	0.67	0.67

4.6 Construct equivalence across language versions

The OPQ32 was first developed in UK English, and localised into different languages as is described in more detail in the OPQ32 Technical Manual (SHL, 2006). To determine whether other language versions of the OPQ32r measure the same constructs as the UK English version, the pattern of scale inter-correlations was examined using structural equation modelling (SEM).

The sample used is sample 6, the International Norm Sample (N = 118,324) which is described in section 2.6. The model tested was whether the 32 OPQ scales had the same pattern of correlations across regions and languages. Both the UK English and US English samples were used as reference languages to which the regional samples were compared. In SEM all inter-correlations in the reference data were constrained to be the same in the regional data and it was then examined how well this constraint fits the actual data. Statistics used to measure how adequately this hypothesized model describes the sample data were the comparative fit index (CFI), which ranges from zero to 1.00, with a value close to or greater than 0.95 indicating a good fit (Hu & Bentler, 1999). Another well-accepted criterion used is the root mean square error of approximation (RMSEA) with values less than 0.05 indicating good fit (Byrne, 2006).

Results of structural equation modelling comparing individual language versions to UK English

The tested models showed excellent fit across regions and languages (see Table 15), with CFI ranging from 0.984 to 0.999 and RMSEA from 0.007 to 0.029 (mean CFI=0.993, mean RMSEA=0.017). Other statistics in Table 15 support this. It should be noted that with large sample sizes, as here, chi-square is nearly always significant and the ratio of chi-square to degrees of freedom does not provide a valid indicator of fit. In conclusion the results strongly confirm the construct (i.e. configural) equivalence of the other OPQ32r language versions with the UK English version as well as across regions.

|

Table 15: Fit statistics comparing UK English with other language regions.

UK English in UK (N=22,612) compared to	N	CFI: good: .95-1.00, acceptable: 0.90-.95				<0.10	<0.05 good model fit, < 0.08 acceptable		
		CFI	χ²	sig	df	SRMR	RMSEA	90% confidence interval RMSEA	
Arabic	1,494	0.990	3440.545	0.00	496	0.078	0.022	0.021	0.023
Belgian Dutch	2,385	0.995	2115.762	0.00	496	0.044	0.016	0.015	0.017
Belgian French	2,529	0.993	2698.000	0.00	496	0.041	0.019	0.018	0.019
Brazilian Portuguese	1,006	0.996	1724.039	0.00	496	0.065	0.014	0.014	0.015
Castilian Spanish	695	0.998	997.268	0.00	496	0.052	0.009	0.008	0.010
Danish	6,809	0.990	4330.408	0.00	496	0.032	0.023	0.022	0.024
Dutch	6,227	0.989	4501.427	0.00	496	0.037	0.024	0.023	0.024
Finnish	5,381	0.989	4455.727	0.00	496	0.045	0.024	0.023	0.025
French	4,225	0.989	4007.578	0.00	496	0.042	0.023	0.022	0.024
German	2,336	0.992	2941.355	0.00	496	0.049	0.020	0.019	0.021
Icelandic	729	0.997	1263.498	0.00	496	0.053	0.012	0.011	0.012
Italian	4,550	0.988	4435.433	0.00	496	0.043	0.024	0.024	0.025
Korean	846	0.994	2129.404	0.00	496	0.061	0.017	0.016	0.017
Latin American Spanish	971	0.997	1353.440	0.00	496	0.06	0.012	0.011	0.013
Norwegian	4,659	0.989	4158.291	0.00	496	0.037	0.023	0.023	0.024
Polish	927	0.998	1148.131	0.00	496	0.041	0.011	0.01	0.011
Portuguese	2,248	0.995	2142.745	0.00	496	0.046	0.016	0.016	0.017
Simplified Chinese	3,322	0.984	5726.448	0.00	496	0.06	0.029	0.028	0.029
Swedish	13,244	0.986	6320.278	0.00	496	0.029	0.026	0.025	0.026
Traditional Chinese	564	0.997	1341.068	0.00	496	0.072	0.012	0.011	0.013
Turkish	1,107	0.995	1973.295	0.00	496	0.06	0.016	0.015	0.017
UK English in Australia	9,120	0.998	1157.460	0.00	496	0.013	0.009	0.008	0.010
UK English in India	1,098	0.997	1464.210	0.00	496	0.051	0.013	0.012	0.014
UK English in Malaysia	2,672	0.992	2862.372	0.00	496	0.042	0.019	0.019	0.020
UK English in New Zealand	2,818	0.999	840.378	0.00	496	0.023	0.007	0.007	0.008
UK English in Singapore	4,053	0.994	2366.683	0.00	496	0.038	0.017	0.016	0.018
UK English in South-Africa	4,880	0.993	2816.973	0.00	496	0.032	0.018	0.018	0.019
US English In US	4,114	0.996	1671.629	0.00	496	0.028	0.013	0.013	0.014
US English in Canada	703	0.999	792.920	0.00	496	0.036	0.007	0.006	0.008

Results of structural equation modelling - comparing individual languages versions to US English

In addition to comparing the language regions to the UK English version, regions were also compared to the US English version as this is a widely used language version of the OPQ. Similar to the UK English results, the tested model showed excellent fit across all language versions (see Table 16), with CFI ranging from 0.960 to 0.996 and RMSEA from 0.013 to 0.044 (mean CFI=0.983, mean RMSEA=0.028). In conclusion the results strongly confirm the construct (i.e. configural) equivalence of the other OPQ32r language versions with the US English version.

Table 16: Fit statistics comparing US English with other language regions.

US English in US (N=4,114) compared to	N	CFI: good: .95-1.00, acceptable: 0.90-.95				<0.10	<0.05 good model fit, < 0.08 acceptable		
		CFI	χ^2	sig	df	SRMR	RMSEA	90% confidence interval RMSEA	
Arabic	1,494	0.963	2855.775	0.00	496	0.063	0.041	0.040	0.043
Belgian Dutch	2,385	0.981	2039.476	0.00	496	0.044	0.031	0.030	0.032
Belgian French	2,529	0.978	2331.652	0.00	496	0.043	0.033	0.032	0.035
Brazilian Portuguese	1,006	0.981	1618.155	0.00	496	0.055	0.030	0.028	0.031
Castilian Spanish	695	0.992	962.738	0.00	496	0.045	0.020	0.018	0.022
Danish	6,809	0.985	2575.093	0.00	496	0.036	0.028	0.027	0.029
Dutch	6,227	0.979	3281.457	0.00	496	0.047	0.033	0.032	0.034
Finnish	5,381	0.982	2723.652	0.00	496	0.046	0.031	0.030	0.032
French	4,225	0.977	2742.342	0.00	496	0.036	0.033	0.032	0.034
German	2,336	0.975	2379.13	0.00	496	0.039	0.034	0.033	0.036
Icelandic	729	0.987	1271.423	0.00	496	0.046	0.025	0.024	0.027
Italian	4,550	0.973	3229.857	0.00	496	0.043	0.036	0.035	0.037
Korean	846	0.976	1983.592	0.00	496	0.060	0.035	0.033	0.036
Latin American Spanish	971	0.987	1283.572	0.00	496	0.052	0.025	0.023	0.027
Norwegian	4,659	0.980	2583.541	0.00	496	0.034	0.031	0.030	0.032
Polish	927	0.988	1212.957	0.00	496	0.045	0.024	0.022	0.026
Portuguese	2,248	0.982	1844.599	0.00	496	0.036	0.029	0.028	0.031
Simplified Chinese	3,322	0.960	4089.779	0.00	496	0.055	0.044	0.043	0.045
Swedish	13,244	0.987	2990.789	0.00	496	0.034	0.024	0.023	0.025
Traditional Chinese	564	0.986	1305.794	0.00	496	0.075	0.026	0.025	0.028
Turkish	1,107	0.979	1849.581	0.00	496	0.056	0.032	0.031	0.034
UK English in Australia	9,120	0.995	1243.029	0.00	496	0.023	0.015	0.014	0.016
UK English in India	1,098	0.985	1411.309	0.00	496	0.052	0.027	0.025	0.028
UK English in Malaysia	2,672	0.979	2167.776	0.00	496	0.036	0.032	0.03	0.033
UK English in New Zealand	2,818	0.993	1070.343	0.00	496	0.026	0.018	0.017	0.020
UK English in Singapore	4,053	0.988	1758.886	0.00	496	0.042	0.025	0.024	0.026
UK English in South-Africa	4,880	0.985	2041.475	0.00	496	0.030	0.026	0.025	0.028
UK English in UK	22,612	0.996	1671.629	0.00	496	0.028	0.013	0.013	0.014
US English in Canada	703	0.995	794.632	0.00	496	0.039	0.016	0.014	0.018

4.7 Relationships of OPQ32r with other instruments

This section provides evidence of the OPQ32r construct validity based on correlations of the OPQ32r with other instruments. Construct validity can be demonstrated by providing convergent evidence (relationships occurring where expected) and divergent evidence (absence of relationships where not expected).

Relationships of OPQ32r scales with the Motivation Questionnaire

Table 17 summarizes the results of three studies exploring relationships between the OPQ32 and the Motivation Questionnaire, or MQ (SHL, 2014). The first two studies, relating to the OPQ32i and OPQ32n used UK English data. The third, OPQ32r, is based on Norwegian data. Because motivation and values are conceptually different from behavioural style as measured in personality questionnaires, these correlations are not expected to be as high as for pairs of concordant personality measures. Therefore all correlations greater than 0.30 or less than -.30 are shown in the table. The full correlation matrices for the UK data are in the OPQ32 Technical Manual (SHL, 2006, Appendix D). The matrix for the Norwegian data is attached to this report as Appendix A.

The first data set of 479 respondents was drawn from the UK general population group (see SHL, 2006, Appendix B) who completed both the OPQ32n and the MQ. The second data set consisted of 353 senior managers who completed the OPQ32i and the MQ as part of a leadership potential development programme. This study is further described in the OPQ32 Technical Manual (Chapter 8). The third data set is an extract of people who have recently completed both the Norwegian language version of OPQ32r and the Norwegian language version of MQ (n=143).

Table 17 lists each OPQ32 scale that correlate with an MQ scale with r>0.30 or r<-.30. For all three data sets correlations between -.30 and 0.30 are omitted. For the senior managers, there was some evidence of range restriction and therefore probable attenuation in correlations. However, the point of Table 17 is to compare the general pattern of relationships. Many smaller correlations not cited here, also reach statistical significance in these large data sets.

Correlations between the two instruments are moderate rather than high. People's self-report of their preferred style of behaving, thinking and feeling is different from the way they describe what motivates them. However, as would be expected there are moderate relationships between scales with similar content domains. For example, OPQ32 Controlling correlates with MQ Power, OPQ32 Competitive with MQ Competition, and OPQ32 Achieving with MQ Achievement.

In general we see that the Norwegian OPQ32r-MQ relationships have similarities with the pattern of results for the English OPQ32n-MQ and OPQ32i-MQ. All three data sets agree on not having any correlations between OPQ32 scales and MQ scales for 11 of the 32 OPQ scales. 14 of the 41 correlations between OPQ32n and MQ are common between OPQ32n and OPQ32r; 8 of the 14 OPQ32i-MQ correlations are common with OPQ32r-MQ. Generally, OPQ Persuasive and Controlling correlate with the MQ Energy scales, as do OPQ Data Rational, Evaluative and Behavioural, reverse Conventional and the OPQ 'dynamism' scales Vigorous through Decisive. OPQ 'Variety Seeking' correlates with the MQ Intrinsic motivation scales. For all three data sets, as one would expect there are relatively few correlations with the MQ Synergy or Extrinsic motivation scales.

Table 17: Correlations between OPQ32 and Motivation Questionnaire.

OPQ32 scale	MQ – OPQ32n UK N=479	MQ – OPQ32i UK N=353	MQ-OPQ32r Norway N=143
Persuasive	Power .45		Commercial Outlook .35 Immersion, .32
Controlling	Power .63 Achievement .40 Interest .36 Personal Growth .34 Level of Activity .31	Power .38	Power .55 Flexibility, .31
Outspoken			
Independent Minded			
Outgoing			
Affiliative	Affiliation .43		
Socially Confident			
Modest			
Democratic	Affiliation .33		
Caring	Affiliation.33	Personal Principles .30 Affiliation 0.30	
Data Rational			Commercial Outlook .33
Evaluative	Achievement .37 Power .35 Interest .42 Personal Growth .31		Power .39
Behavioural	Personal Principles .31 Personal Growth .31		Flexibility .38 Power .34
Conventional	Interest –.36 Flexibility –.38 Power -.31	Flexibility –.41	Interest -.34 Flexibility -.37 Power -.31
Conceptual	Personal Growth .37 Interest .39 Achievement .33		
Innovative	Power .31 Interest .33		
Variety Seeking	Interest .46 Achievement .32 Power .33 Personal Growth .31 Flexibility .33		Power .36 Interest .48 Personal Growth .32 Flexibility, .35
Adaptable			
Forward Thinking	Power .34 Personal Growth .33		Power .31
Detail Conscious		Flexibility –.37	
Conscientious			Flexibility -.30
Rule Following	Flexibility –.39	Flexibility –.55	Flexibility -.38
Relaxed			
Worrying			Power -.34
Tough Minded			
Optimistic			
Trusting			
Emotionally Controlled			
Vigorous	Level of Activity .43 Achievement .35 Power .31	Level of Activity .35	Level of Activity 0.35
Competitive		Competition .52 Commercial Outlook .30	Competition .47 Commercial Outlook 0.37
Achieving	Achievement .43 Power .51 Personal Growth .41 Interest .35 Progression .34	Achievement .33 Competition .35 Progression .42	Level of activity .32 Achievement .33 Competition .39 Power .36 Immersion.35
Decisive	Power .32		Flexibility .32

UK OPQ32i and OPQ32n data (from the OPQ32 Technical Manual, 2006) and Norwegian OPQ32r data from 2010-2011;

Only correlations greater than 0.30 or less than -0.3 are shown. All correlations are significant at the 0.01 level (2-tailed).

Concordance analysis of OPQ-MQ relationships

Research carried out in 2006 examined the levels of concordance between item content on OPQ32 and MQ. On the basis of judged conceptual concordance a matrix was created where the expected relationship between each OPQ32 scale and each MQ scale was coded as follows: 0 if there was expected to be no relationship; +1 or -1 if a weak relationship (positive or negative) was expected and +2 or -2 if a strong positive or negative relationship was expected. There were two subject matter experts, both of whom knew the content of both instruments very well. The judges defined a 'weak' relationship as implying an expected correlation of around 0.20 and a 'strong' one as around 0.40. Agreement on the concordances was reached prior to examining the fit between the concordance matrix and the correlation matrices. (Details of the rationale and procedures for producing concordance matrices are presented in Warr, 1999).

The mean and SD of the observed correlations for each of the hypothesized levels of concordance was then computed as was the overall correlation between the observed 576 correlations (32 OPQ scales by 18 MQ scales) and the 576 expected concordances. Average correlations for the hypothesized zero correlations were computed both for the absolute values of the correlations and for actual values. The results of the original 2006 analyses for OPQ32i and OPQ32n, together with the results for the 2011 Norwegian OPQ32r data (using the same 2006 concordance table) are shown in Table 18.

It can be seen that the results for the Norwegian OPQ32r data are much closer to the original UK OPQ32n results than the OPQ32i results. The range of correlation means from concordance=+2 to concordance=-2 for OPQ32n was 0.54. For OPQ32i this was 0.13 while for the Norwegian OPQ32r data it is 0.49. The overall correlations (N=576) between the obtained correlations and predicted concordance values was r=0.55 for OPQ32n, 0.26 for OPQ32i and 0.50 for the Norwegian OPQ32r data.

Table 18: Average observed correlates for each level of concordance with SDs and average absolute value for zero-level concordance.

Concordance	OPQ32n, UK data N=479		OPQ32i UK data N=353		OPQ32r Norwegian data N=143		N of correlations
	Mean	SD	Mean	SD	Mean	SD	
+2	0.34	0.11	0.11	0.20	0.26	0.14	19
+1	0.21	0.11	0.08	0.12	0.16	0.12	42
0	0.08	0.12	-.01	0.10	0.02	0.12	481
-1	-.11	0.13	-.06	0.11	-.09	0.12	29
-2	-.19	0.17	-.02	0.16	-.23	0.10	9
ABS(r) for 0	0.11	0.08	0.08	0.07	0.10	0.08	481

All comparisons use the same 2006 concordance matrix.

This confirms the impression that the Norwegian OPQ32r data produces good evidence of construct validity that is stronger than that from the UK OPQ32i and that is more in line with that from the UK OPQ32n data set.

Relationships of OPQ32r scales with the Hogan Development Survey (HDS)

This study is based on data collected with 510 volunteers who had completed the HDS (Hogan & Hogan, 1997) and in addition were asked to complete the OPQ32r specifically for this research study. All participants were employed at the time of data collection. Of the assessment takers 63% were male and 37% female. Information regarding age was recorded for 82% of the group, with the majority falling into the following age groups: 21-24: 2%, 25-29: 5%, 30-34: 12%, 35-39: 24%, 40-44: 22%, 45-49: 22%, 50-54: 9%, 55-59: 4% and 60-64: 1%.

The HDS assesses "dark side traits" in a normal population and has been used in various studies to investigate dysfunctional behaviour at work. The HDS is not a clinical measure, but rather measures dysfunctional personality in the working population using a similar taxonomy to the classical personality disorders. An overview of the HDS scales is provided below:

- **Excitable:** moody, easily annoyed, hard to please, and emotionally volatile
- **Sceptical:** distrustful, cynical, sensitive to criticism, and focused on the negative
- **Cautious:** fearful of criticism and making mistakes, easily embarrassed and unwilling to take chances
- **Reserved:** aloof, indifferent to the feelings of others, and uncommunicative
- **Leisurely:** overtly cooperative, but privately irritable, stubborn, and uncooperative

- **Bold:** overly self-confident, arrogant, with inflated feelings of self-worth

- **Mischievous:** charming, risk-taking, limit-testing and excitement-seeking

- **Colourful:** dramatic, attention-seeking, interruptive, and poor listening skills

- **Imaginative:** creative, but thinking and acting in unusual or eccentric ways

- **Diligent:** meticulous, precise, hard to please, and tends to micromanage

- **Dutiful:** eager to please and reluctant to act independently or against popular opinion

Correlations between the OPQ and HDS dimensions are presented in Table 19. HDS and OPQ scales of conceptually similar constructs correlated in the expected direction, with correlations being moderately high. Overall, the study indicates that many of the constructs measured in both instruments, assessing positive and maladaptive aspects of personality, share much variance. This study helps to understand the overlap between desirable and maladaptive personality traits and further supports the construct validity of the OPQ32r. We thank PCL for providing the HDS data and Gillian Hyde for making this research collaboration possible.

Table 19: Correlations between OPQ and HDS scales (N=510).

OPQ32r scales	HDS EXC	HDS SCE	HDS CAU	HDS RES	HDS LEI	HDS BOL	HDS MISCH	HDS COL	HDS IMA	HDS DIL	HDS DUT
Persuasive	-.241	-.031	-.513	-.277	-.165	.181	.273	.503	.153	-.076	-.149
Controlling	-.224	.028	-.451	-.214	-.135	.331	.245	.454	.083	-.069	-.265
Outspoken	.009	.112	-.319	-.081	-.212	.203	.163	.263	.173	-.121	-.306
Independent Minded	.113	.190	-.031	.206	.108	.161	.153	.107	.283	-.099	-.372
Outgoing	.004	.088	-.339	-.346	-.070	.192	.263	.533	.165	-.202	-.073
Affiliative	-.141	-.128	-.181	-.414	-.153	.090	.115	.340	-.030	-.139	.071
Socially Confident	-.286	-.036	-.550	-.420	-.179	.090	.249	.447	.082	-.099	-.129
Modest	.110	-.057	.300	.230	.074	-.267	-.203	-.344	-.114	.045	.145
Democratic	-.108	-.210	-.129	-.216	-.215	-.105	-.088	.050	-.072	-.032	.194
Caring	-.117	-.207	-.021	-.290	-.153	-.048	-.061	.065	.025	-.017	.208
Data Rational	-.152	-.085	-.171	.005	.014	.119	-.003	-.046	.005	.125	-.073
Evaluative	-.021	-.007	-.181	.093	-.072	.099	-.053	-.030	-.001	.100	-.115
Behavioural	-.105	-.086	-.175	-.194	-.124	.074	.103	.207	.157	-.136	.057
Conventional	.052	.011	.239	.054	.027	-.161	-.284	-.300	-.365	.311	.237
Conceptual	-.033	-.014	-.187	-.050	-.040	.127	.124	.134	.258	-.089	-.100
Innovative	-.143	-.023	-.298	-.116	-.067	.199	.242	.265	.308	-.141	-.169
Variety Seeking	.120	.033	-.032	.056	.015	.110	.254	.211	.203	-.340	-.185
Adaptable	.161	.034	.111	.019	.061	-.035	.028	.034	.047	-.039	.083
Forward Thinking	-.243	-.052	-.303	-.131	-.116	.176	.013	.152	.122	.085	-.121
Detail Conscious	-.149	-.026	-.053	-.040	-.066	-.034	-.211	-.209	-.177	.499	.086
Conscientious	-.182	-.064	-.050	-.071	-.066	-.013	-.204	-.158	-.201	.397	.108
Rule Following	-.033	-.033	.122	-.057	-.002	-.079	-.319	-.222	-.291	.391	.236
Relaxed	-.319	-.162	-.267	-.104	-.128	.022	.098	.084	-.069	-.118	-.076
Worrying	.292	.003	.509	.229	.150	-.220	-.250	-.350	-.118	.043	.242
Tough Minded	-.272	-.124	-.334	-.063	-.084	.007	.047	.088	.029	-.071	-.106
Optimistic	-.417	-.221	-.386	-.329	-.130	.143	.194	.331	.075	-.130	-.082
Trusting	-.225	-.427	-.115	-.178	-.207	-.045	-.066	.047	-.093	-.146	.099
Emotionally Controlled	.038	.093	.245	.341	.281	-.065	-.091	-.300	-.094	.058	.062
Vigorous	-.108	-.091	-.057	-.118	-.017	-.034	-.024	.034	-.084	.136	.090
Competitive	-.017	.198	-.246	-.077	.032	.259	.289	.308	.086	.032	-.253
Achieving	-.254	-.020	-.466	-.281	-.088	.324	.217	.362	.173	.065	-.259
Decisive	-.165	-.013	-.277	-.058	-.101	.264	.178	.246	.084	-.203	-.228

All absolute correlations >0.086 are significant at p< 0.05; All absolute correlations >0.115 are significant at p <0.01.

Relationships of OPQ32r with tests of cognitive ability

In general we do not expect to find correlations between tests of ability and measures of personality. The OPQ32 Technical Manual (2006) reports an exception to this general rule. People who described themselves as high on 'Data Rational' on the OPQ32 also tend to have high ability test scores. There is also a tendency for there to be a relationship between ability level and scores on 'Evaluative'. Some negative correlations with ability were also found. The results are summarised as follows:

"Results across different data sets confirm that the OPQ32 measures aspects of people that are generally independent of ability. However, there are some consistent but moderate correlations between ability test performance and a limited number of personality traits. In all data sets a significant correlation was found between the OPQ32 scale Data Rational and scores on different numerical ability tests. As outlined above, that is expected from the fact that individuals who are good at numerical reasoning are more likely to enjoy working with numbers.

There were also consistent tendencies for certain ability scales to have small but significant negative correlations with some aspects of personality: Outgoing, Persuasive, Caring and Affiliative all had significant negative correlations with ability in more than one data set." (SHL, 2006).

For the Norwegian version of Verify Verbal and Verify Numerical there are data sets of N=189 and N=175, respectively for people with data on both the ability test and the OPQ32r. The correlations of the OPQ32 scales with ability are shown in Table 20. As in previous data sets we find the strongest positive correlations with Data Rational for Numerical Ability and with both ability tests for Evaluative. There are some additional significant correlations showing a similar pattern to previous results.

Table 20: Correlations between OPQ32r and Verify Managerial Professional Numerical (N=175) and Verbal (N=189) Reasoning tests (all in Norwegian).

OPQ32r scale	Numerical	Verbal
Persuasive	-.115	-.011
Controlling	**-.212**[**]	-.014
Outspoken	.068	.082
Independent Minded	.041	-.049
Outgoing	-.094	.012
Affiliative	-.032	.046
Socially Confident	-.043	-.065
Modest	.077	.011
Democratic	-.038	-.071
Caring	-.139	.035
Data Rational	**.342**[**]	.060
Evaluative	**.214**[**]	**.217**[**]
Behavioural	-.036	.106
Conventional	-.162[*]	-.068
Conceptual	.159[*]	.008
Innovative	-.083	-.037
Variety Seeking	.119	.037
Adaptable	.006	-.011
Forward Thinking	.001	.133
Detail Conscious	-.014	-.090
Conscientious	-.047	-.014
Rule Following	-.005	-.082
Relaxed	.099	.085
Worrying	-.096	.031
Tough Minded	**.215**[**]	.136
Optimistic	-.095	.050
Trusting	-.057	.071
Emotionally Controlled	.015	**-.200**[**]
Vigorous	-.008	-.069
Competitive	.056	.069
Achieving	.080	.111
Decisive	-.040	.092

[**] p<.01 also highlighted in bold print.

In summary, the OPQ32r scale correlations represent the correlations between OPQ32 constructs, without the distortion caused by ipsative constraints. They are no longer lower than the respective correlations of the OPQ32n.

Construct relationships of the full OPQ32i with its sibling measure, OPQ32n, are preserved in the IRT-scored shortened version. The relationships between OPQ32r and the OPQ32n are strong. Just as for the OPQ32n, six or seven factors are typically extracted from the IRT-scored OPQ32r. The main six factors represent the Big Five factors of personality with Openness split into two facets, and additional factors can vary across samples.

Results of structural equation modelling comparing language versions of the OPQ32r strongly confirm the construct (i.e. configural) equivalence of the other OPQ32r language versions with the UK and US English versions.

Comparisons with other instruments measuring motivation, maladaptive personality traits and cognitive ability showed high correlations between scales with similar content and low correlations between scales that are different.

In addition to the evidence directly associated with OPQ32r in this section, a large body of evidence has been assembled for the construct validity of OPQ32n and OPQ32i in the OPQ32 Technical Manual (SHL, 2006). The three instruments were developed under the same concept model and are functioning in similar ways. In particular, as noted in the previous sections, the three instruments have similar scale covariance structures, the scales group into similar factors when factor analysed and the instruments produce highly correlated scores in test-retest studies. Therefore, there is reason to support the transportability of construct validity evidence to OPQ32r from evidence associated with OPQ32n and OPQ32i.

Chapter 5: Criterion-related validity

Criterion-related validity describes relationships between scores on a psychometric measure and external criteria. For the OPQ32 with its work-related personality dimensions, such criteria are typically indicators of performance of various kinds. When measuring performance of an employee within a particular role, organisations can utilise three types of data: judgemental (such as line-manager ratings of competencies), production (such as sales performance data) and personnel (such as absenteeism records) data. Our approach to summarising validation studies has been to utilise the Universal Competency Framework as a common structure for describing judgemental criterion data across studies (see Bartram, 2005). Validation studies presented in the 2006 OPQ32 Technical Manual all use the UCF 20 Dimensions as the common model for mapping competency models used in specific validation studies and as the model for mapping predictor scales. The reader is referred to the 2006 Manual and Bartram (2005) for more details of the rationale and the methodology.

5.1 Criterion validity evidence from applying the IRT scoring model to OPQ32i data

As demonstrated earlier, the IRT-based scores of the OPQ32r correlate strongly with its predecessor, OPQ32i. It has also been shown that the IRT scoring preserves the correlations with external measures, such as the OPQ32n. The question is whether the IRT scoring preserves the well-established criterion-related validity of OPQ32i. In other words, is the variance predictive of job performance and other outcomes captured in the IRT-based scores of the reduced number of items of the OPQ32i? If so, criterion validity findings for the OPQ32i can be generalised to the OPQ32r. To answer this question, a large OPQ32i validation study will be examined. The validities of two types of scores will be compared – the ipsative OPQ32i scores and the IRT-scored responses to the OPQ32i items that make the OPQ32r.

OPQ32r as a predictor of management competence in development settings

This concurrent study was carried out in an organisation in the food manufacturing industry to identify personality constructs that could facilitate developmental opportunities amongst Directors and Senior Managers across different countries. The original study was presented in the OPQ32 Technical Manual (SHL, 2006; chapter 8, study 2.1). Since 2005, the sample size has doubled and the present investigation includes this new data.

Validation study of managerial competencies (N=853). Eight hundred and fifty three Directors and Senior Managers from the company took part. They were located across Europe, Asia Pacific, North, Central and South America. Eighty-one percent were male. Ages ranged between 35 and 60 years, with the majority of participants being French nationals and almost all educated to university level.

Instruments. A concurrent validity study was conducted using the OPQ32i as a predictor of performance. The appropriate language version of the OPQ32i was used in different countries (English, French, German, Italian, Portuguese, and Spanish).

The Inventory of Management Competencies (IMC) was used as the 360-degree tool to obtain performance ratings. The IMC was completed by self, manager/s, colleague/s and direct report/s, in the appropriate language version.

Of the feedback collected:

- 49.6% had two or more manager ratings (range 1–4, mean=1.54, SD=0.59).

- 98.6% had two or more direct report ratings (range 0–10, mean=3.24, SD=1.17).

- 98.4% had two or more colleague ratings (range 1–15, mean=3.21, SD=1.28).

Descriptions of the IMC competencies and their mappings to the Universal Competency Framework (UCF) model are provided for reference. The mappings show the *closest* competencies from the UCF model, but the competency definitions sometimes differ. For example, IMC 'Action Orientation' and UCF 1.1 'Deciding and Initiating Action' cover similar, but not the same, areas of competence. The UCF competency includes making tough decisions while the IMC competency does not.

Method. The validation methodology described in chapter 8 (Criterion-related Validity) of the OPQ32 Technical Manual (SHL, 2006) was applied to the data obtained from the IMC and OPQ32i. This is a theory-driven approach, which is commonly described as "criterion-centric" (Bartram, 2005; Bartram, Warr & Brown, 2010).

- **Average performance ratings by category.** Ratings by different people within each rater category were averaged to produce average category ratings.

- **Correlations between composite prediction scores and average performance ratings by rater category.** Composite OPQ32 prediction scores derived from the OPQ32 scales were produced for each of the competencies. The composite scores were derived from the OPQ32 traits that are hypothesised to be predictive of certain types of behaviour, relevant to the context of this study (see Table 22). Such scales were identified *a priori* for each IMC competency and then summed (no weighting was applied) to produce the composite score. The composite scores were correlated with actual performance ratings, separately for each rater category.

- **Average correlations for non-hypothesised scales.** Correlations with all other OPQ32 scales, hypothesised to be unrelated to the behaviour, were averaged. The direction of the scale was taken into account, and scales that negatively related to the Big Five were reversed in order to provide consistent direction of relationships.

- **Correlations between the Big Five and average performance ratings by rater category.** In addition, composite Big Five scores were produced from OPQ32 scales, following the mappings described in the OPQ32 Technical Manual (SHL, 2006; chapter 7 on Construct Validity), and then correlated with each of the competencies.

- **Median validity of composite personality predictors by rater category.** The median was computed from all correlations between performance ratings for each rater category and composite personality predictors.

In every analysis, comparisons were made between the OPQ32r (i.e. IRT-scored OPQ32i responses to items that formed OPQ32r) and the ipsative OPQ32i. Two composite prediction scores were produced for each of the competencies, based on these two sets of scores. The composite Big Five scores were also produced based on the OPQ32r and the OPQ32i.

Results and Discussion. Table 22 shows the correlations between the composite OPQ32 predictors (based on both the OPQ32r and the OPQ32i scores) and the performance ratings on the 16 IMC competencies by rater category. The correlations are not corrected for any artefacts. The table also includes the average correlations between non-hypothesised OPQ32 scales and the performance ratings. For all of the rater groups, results are in line with the hypotheses. As expected, self-ratings show the strongest relationships with the personality predictors; the median correlation of composite personality predictors is 0.33 for the OPQ32r and 0.35 for the OPQ32i. Best composite validities for OPQ32 predictions reach as high as 0.29 (IRT-scored OPQ32r) and 0.30 (ipsative OPQ32i) for manager, 0.30 (OPQ32r) and 0.33 (OPQ32i) for colleagues, 0.27 (OPQ32r) and 0.30 (OPQ32i) for direct reports. For readily observable competencies such as Action Orientation, Personal Motivation, Leadership, Creativity and Innovation, Interpersonal Sensitivity and Persuasiveness, validities are high for all rater categories. For competencies that are less visible to others, such as Resilience (Funder & Dobroth, 1987), self-ratings have much higher validities than the ratings by others. Both the OPQ32r and the OPQ32i show that personality scales have generally low validities in predicting technical competencies such as Specialist Knowledge, Written Communication, and Problem Solving and Analysis. This is in line with other research findings. Ability tests are much more relevant and valid predictors in areas of technical competence.

Validities of composite personality predictors based on the OPQ32r in relation to the 16 different competencies as assessed by managers, colleagues and direct reports are very close to the validities for the composite predictors based on the ipsative OPQ32i (none of differences in correlations are statistically significant). This confirms that the shorter IRT-scored version preserved the validity of the full OPQ32i.

Table 21: Definitions of the IMC competencies, their relationship to the 20 UCF competencies and hypothesised relationships to the OPQ32 scales.

IMC Competency	Description	Hypothesised OPQ32 scales	Mapping to UCF 20
Action Orientation	Demonstrates a readiness to make decisions, take the initiative and originate action.	Decisive, Controlling, Outspoken, NOT Democratic, NOT Worrying	UCF 1.1
Leadership	Motivates and empowers others in order to reach organisational goals.	Controlling, Caring, NOT Independent Minded, NOT Emotionally Controlled	UCF 1.2
Interpersonal Sensitivity	Interacts with others in a sensitive and effective way. Respects and works well with others.	Democratic, Caring, NOT Competitive, NOT Outspoken	UCF 2.1
Persuasiveness	Influences, convinces or impresses others in a way that results in acceptance, agreement or behaviour change.	Persuasive, Socially Confident, NOT Worrying, NOT Emotionally Controlled	UCF 3.2
Oral Communication	Speaks clearly, fluently and in a compelling manner to both individuals and groups.	Persuasive, Socially Confident, Adaptable	UCF 3.3
Written Communication	Writes in a clear and concise manner, using appropriate grammar, style and language for the reader.	Persuasive, Socially Confident, Detail Conscious	UCF 4.1
Specialist Knowledge	Understands technical or professional aspects of work and continually maintains technical knowledge.	Data Rational, NOT Conventional	UCF 4.2
Problem Solving and Analysis	Analyzes issues and breaks them down into their component parts. Makes systematic and rational judgements based on relevant information.	Data Rational, Evaluative	UCF 4.3
Creativity and Innovation	Creates new and imaginative approaches to work-related issues. Identifies fresh approaches and shows a willingness to question traditional assumptions.	Innovative, NOT Conventional	UCF 5.2
Strategic	Demonstrates a broad-based view of issues, events and activities and a perception of their longer term impact or wider implications.	Forward Thinking, Achieving, NOT Conventional	UCF 5.3
Planning and Organising	Organises and schedules events, activities and resources. Sets up and monitors timescales and plans.	Forward Thinking, Detail Conscious, Conscientious, Controlling	UCF 6.1
Quality Orientation	Shows awareness of goals and standards. Follows through to ensure that quality and productivity standards are met.	Detail Conscious, Achieving	UCF 6.2
Flexibility	Successfully adapts to changing demands and conditions.	Adaptable, Optimistic, NOT Independent Minded	UCF 7.1
Resilience	Maintains effective work behaviour in the face of setbacks or pressure. Remains calm, stable and in control of themselves.	Relaxed, Tough Minded, Optimistic, Emotionally Controlled, NOT Independent Minded	UCF 7.2
Personal Motivation	Commits self to work hard towards goals. Shows enthusiasm and career commitment.	Achieving, Vigorous, Controlling, NOT Emotionally Controlled	UCF 8.1
Commercial Awareness	Understands and applies commercial and financial principles. Views issues in terms of costs, profits, markets and added value.	Competitive, Achieving, Persuasive, Data Rational	UCF 8.2

Table 23 shows correlations between composite Big Five scores (based on both OPQ32r and OPQ32i) and performance by rater category. Extraversion is a strong predictor of Persuasiveness and Oral Communication (as assessed by all rater categories); Openness to Experience was a strong predictor of Creativity and Innovation (for all rater categories); Agreeableness predicted IMC Interpersonal Sensitivity (all rater categories); and Conscientiousness predicted Planning and Organising together with Quality Orientation (most rater categories). Emotional Stability predicted Resilience, where correlations were strong for self-ratings and weak for all other rater categories. Weak correlations between IMC Resilience and Emotional Stability for non-self ratings are likely to be due to the low observability of this behaviour. Self-assessment in this case can be particularly useful.

There are insignificant differences between validity coefficients for the IRT-scored OPQ32r and the ipsative OPQ32i for all but three cases. The IRT-scored OPQ yields significant increases in validity for Extraversion in relation to the competency Personal Motivation and for Conscientiousness in relation to the competencies Action Orientation and Personal Motivation. These increases are due to removing ipsative constraints from correlations between OPQ primary scales, and thus increasing the variance in composite scores consisting of several correlated traits (as in deriving Big Five from narrow OPQ traits).

Table 22: Correlations between composite OPQ32 predictors and performance ratings by rater category.

IMC Competency / Mapping to the UCF			Action Orientation UCF 1.1	Leadership UCF 1.2	Interpersonal Sensitivity UCF 2.1	Persuasiveness UCF 3.2	Oral Communication UCF 3.3	Written Communication UCF 4.1	Specialist Knowledge UCF 4.2	Problem solving and Analysis UCF 4.3	Creativity and Innovation UCF 5.2	Strategic UCF 5.3	Planning and Organising UCF 6.1	Quality Orientation UCF 6.2	Flexibility UCF 7.1	Resilience UCF 7.2	Personal Motivation UCF 8.1	Commercial Awareness UCF 8.2	Median
Composite predictor																			
Self (N=853)	IRT		**0.43**	**0.31**	**0.33**	**0.39**	**0.27**	**0.17**	**0.12**	**0.24**	**0.56**	**0.31**	**0.43**	**0.32**	**0.22**	**0.36**	**0.48**	**0.39**	0.33
	Ipsative		**0.42**	**0.36**	**0.35**	**0.41**	**0.23**	**0.18**	**0.10**	**0.21**	**0.60**	**0.34**	**0.44**	**0.36**	**0.24**	**0.34**	**0.49**	**0.38**	0.35
Manager (N=853)	IRT		**0.24**	**0.16**	**0.21**	**0.20**	**0.13**	0.00	**0.13**	**0.11**	**0.26**	**0.13**	**0.11**	**0.11**	**0.10**	**0.17**	**0.29**	**0.24**	0.15
	Ipsative		**0.22**	**0.22**	**0.25**	**0.23**	**0.13**	-0.01	**0.12**	**0.13**	**0.27**	**0.14**	**0.13**	**0.13**	**0.14**	**0.19**	**0.30**	**0.24**	0.17
Peers (N=853)	IRT		**0.23**	**0.18**	**0.22**	**0.18**	**0.15**	0.01	0.06	0.06	**0.26**	0.05	**0.12**	**0.08**	**0.11**	**0.14**	**0.30**	**0.24**	0.15
	Ipsative		**0.24**	**0.22**	**0.24**	**0.22**	**0.15**	0.02	*0.08*	*0.07*	**0.27**	*0.07*	**0.18**	**0.12**	**0.12**	**0.15**	**0.33**	**0.25**	0.16
Subordinates (N=853)	IRT		**0.24**	**0.14**	**0.16**	**0.14**	0.03	0.00	0.04	0.07	**0.27**	0.03	**0.15**	**0.10**	0.01	**0.10**	**0.24**	**0.15**	0.12
	Ipsative		**0.23**	**0.19**	**0.17**	**0.15**	0.00	0.01	0.05	0.06	**0.30**	0.04	**0.20**	**0.13**	0.04	**0.13**	**0.27**	**0.16**	0.14
Average non-hypothesised correlation																			
Self (N=853)	IRT		0.08	0.08	0.01	0.08	0.09	0.04	0.05	0.04	0.08	0.06	0.04	0.02	0.05	0.05	0.08	0.06	0.06
	Ipsative		0.02	0.02	0.00	0.02	0.05	0.02	0.02	0.01	0.03	0.02	0.01	0.00	0.03	0.02	0.02	0.02	0.02
Manager (N=853)	IRT		0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.02
	Ipsative		0.02	0.02	0.00	0.02	0.02	0.01	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.01
Peers (N=853)	IRT		0.00	0.00	0.00	0.00	0.00	0.01	0.00	-0.01	0.01	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.01
	Ipsative		0.01	0.00	-0.01	0.01	0.02	0.00	0.00	0.00	0.02	0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.00
Subordinates (N=853)	IRT		0.02	0.01	-0.02	0.02	0.03	0.02	0.01	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.03	0.01	0.01
	Ipsative		0.01	0.00	-0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.01

Bold – correlation significant at the 0.01 level (one-tailed)
Bold Italic – correlation significant at the 0.05 level (two-tailed)

Table 23: Correlations between the Big Five scores derived from OPQ32 and performance ratings by rater category

		Action Orientation	Leadership	Interpersonal Sensitivity	Persuasiveness	Oral Communication	Written Communication	Specialist Knowledge	Problem solving and Analysis	Creativity and Innovation	Strategic	Planning and Organising	Quality Orientation	Flexibility	Resilience	Personal Motivation	Commercial Awareness
SELF (N=853)																	
Extraversion	IRT	**0.24**	**0.24**	0.03	**0.34**	**0.31**	*0.07*	**0.15**	0.03	**0.26**	**0.14**	0.05	0.03	0.05	0.03	**0.28**	**0.20**
	Ipsative	**0.16**	**0.18**	0.01	**0.31**	**0.30**	0.03	**0.11**	-0.03	**0.19**	0.07	-0.02	-0.03	0.01	-0.05	**0.19**	**0.13**
Openness to	IRT	**0.17**	*0.09*	-0.03	**0.17**	**0.15**	0.05	**0.09**	**0.09**	**0.46**	**0.20**	-0.06	-0.04	**0.11**	0.05	**0.16**	*0.07*
	Ipsative	**0.09**	0.00	-0.07	0.04	0.05	-0.01	0.04	0.04	**0.48**	**0.15**	**-0.13**	*-.07*	**0.10**	-0.03	*0.07*	0.00
Agreeableness	IRT	-0.05	**0.10**	**0.31**	0.03	*0.09*	0.02	-0.06	-0.06	0.00	0.02	0.04	0.00	**0.14**	0.04	-0.01	**-0.11**
	Ipsative	-0.06	*0.07*	**0.35**	-0.01	0.05	-0.02	**-0.10**	**-0.11**	-0.04	-0.02	0.02	-0.03	**0.15**	0.04	*-0.08*	**-0.16**
Emotional Stability	IRT	**0.15**	**0.11**	**0.10**	**0.18**	**0.22**	**0.14**	*0.08*	**0.09**	**0.15**	*0.08*	**0.11**	0.02	**0.17**	**0.33**	**0.14**	**0.13**
	Ipsative	**0.09**	0.05	**0.12**	**0.15**	**0.21**	**0.15**	0.05	*.08*	**0.11**	0.06	0.06	-0.02	**0.18**	**0.34**	0.05	*0.08*
Conscientiousness	IRT	**0.24**	**0.22**	-0.04	**0.11**	*0.09*	**0.13**	**0.20**	**0.18**	*0.09*	**0.16**	**0.43**	**0.32**	0.00	**0.15**	**0.34**	**0.25**
	Ipsative	**0.14**	**0.14**	-0.04	-0.02	-0.02	**0.11**	**0.16**	**0.14**	-0.04	**0.11**	**0.42**	**0.33**	-0.03	**0.10**	**0.26**	**0.17**
MANAGER (N=853)																	
Extraversion	IRT	**0.16**	**0.16**	-0.01	**0.21**	**0.18**	0.02	**0.11**	-0.02	**0.17**	*0.08*	-0.03	-0.01	0.00	-0.05	**0.21**	**0.13**
	Ipsative	**0.15**	**0.16**	0.03	**0.23**	**0.20**	0.01	**0.12**	-0.01	**0.17**	*0.08*	-0.03	-0.01	0.03	-0.04	**0.20**	**0.11**
Openness to	IRT	**0.11**	0.06	-0.06	**0.11**	*0.09*	0.03	0.06	0.02	**0.23**	**0.11**	-0.04	-0.03	-0.01	-0.02	**0.13**	0.04
	Ipsative	*0.07*	0.00	-0.06	0.03	0.01	-0.01	0.03	0.01	**0.24**	*0.08*	-0.06	-0.02	-0.01	-0.04	*0.08*	0.00
Agreeableness	IRT	**-0.09**	0.05	**0.18**	0.02	0.02	0.01	-0.03	-0.04	-0.03	0.00	0.01	0.04	0.02	0.04	*-0.08*	**-0.09**
	Ipsative	-0.06	*0.07*	**0.23**	0.01	0.00	-0.03	-0.02	-0.03	-0.03	0.00	0.04	0.07	0.05	*0.08*	**-0.09**	**-0.09**
Emotional Stability	IRT	0.06	*0.07*	0.04	**0.11**	**0.11**	0.05	0.06	0.02	0.03	0.04	0.04	0.02	0.02	**0.09**	0.05	*0.07*
	Ipsative	0.04	*0.07*	**0.09**	**0.12**	**0.11**	0.06	0.05	0.04	0.03	0.05	0.04	0.02	0.06	**0.13**	0.01	0.05
Conscientiousness	IRT	0.07	0.02	**-0.12**	-0.04	-0.01	0.02	0.00	-0.03	-0.02	-0.02	**0.10**	0.06	-0.04	-0.06	**0.13**	0.05
	Ipsative	0.00	-0.05	**-0.12**	**-0.13**	**-0.10**	-0.03	-0.05	-0.04	-0.06	*-0.07*	**0.10**	*0.07*	-0.03	-0.06	0.07	-0.01

Bold – correlation significant at the 0.01 level (two-tailed)
***Bold Italic** –* correlation significant at the 0.05 level (two-tailed)

Table 23 continued

		Action Orientation	Leadership	Interpersonal Sensitivity	Persuasiveness	Oral Communication	Written Communication	Specialist Knowledge	Problem solving and Analysis	Creativity and Innovation	Strategic	Planning and Organising	Quality Orientation	Flexibility	Resilience	Personal Motivation	Commercial Awareness
PEERS (N=853)																	
Extraversion	IRT	**0.17**	**0.13**	0.00	**0.19**	**0.19**	0.01	**0.11**	-0.02	**0.16**	*0.07*	-0.03	-0.02	0.03	-0.04	**0.21**	**0.11**
	Ipsative	**0.17**	**0.14**	0.03	**0.21**	**0.21**	0.02	**0.10**	0.00	**0.16**	*0.08*	-0.04	-0.01	0.04	-0.03	**0.21**	**0.09**
Openness to Experience	IRT	*0.07*	-0.03	*-0.09*	*0.07*	*0.07*	-0.04	0.03	-0.04	**0.23**	0.04	**-0.11**	*-0.08*	0.01	-0.03	**0.10**	-0.01
	Ipsative	0.05	-0.05	*-0.07*	0.03	0.03	-0.05	0.02	-0.04	**0.22**	0.03	**-0.10**	-0.07	0.02	-0.03	*0.07*	-0.02
Agreeableness	IRT	-0.05	*0.07*	**0.21**	0.04	*0.08*	0.04	-0.03	0.01	0.00	0.01	0.02	0.06	*0.08*	0.01	-0.04	*-0.07*
	Ipsative	-0.04	*0.08*	**0.24**	0.03	0.05	0.00	-0.04	0.01	0.01	0.01	0.02	0.06	**0.10**	0.05	-0.07	*-0.09*
Emotional Stability	IRT	0.01	-0.02	-0.02	0.06	*0.08*	0.06	0.03	-0.02	0.04	-0.02	0.01	*-0.08*	0.02	0.05	-0.03	0.01
	Ipsative	-0.02	-0.03	0.02	0.06	**0.09**	*0.07*	0.01	-0.01	0.04	-0.02	-0.01	**-0.09**	0.03	*0.07*	*-0.08*	-0.03
Conscientiousness	IRT	0.03	-0.01	**-0.15**	-0.03	-0.03	0.02	-0.03	-0.03	-0.02	-0.01	**0.11**	*0.07*	*-0.07*	*-0.07*	**0.13**	0.05
	Ipsative	-0.02	-0.02	**-0.11**	**-0.10**	**-0.09**	0.01	-0.06	-0.02	-0.07	-0.02	**0.15**	**0.12**	*-0.07*	*-0.07*	*0.08*	-0.01
SUBORDINATES (N=853)																	
Extraversion	IRT	**0.14**	**0.09**	0.00	**0.15**	**0.14**	0.01	0.06	0.02	**0.16**	0.06	-0.02	0.01	0.02	-0.02	**0.16**	**0.10**
	Ipsative	**0.13**	**0.09**	0.01	**0.17**	**0.16**	0.01	0.06	0.01	**0.14**	0.05	-0.03	0.01	0.02	-0.03	**0.13**	**0.09**
Openness to Experience	IRT	*0.08*	0.00	-0.02	0.06	0.04	-0.04	0.03	0.02	**0.21**	*0.07*	*-0.08*	-0.04	0.04	0.04	**0.10**	-0.01
	Ipsative	0.07	-0.02	-0.01	0.05	0.02	-0.04	0.03	0.02	**0.23**	0.06	*-0.09*	-0.04	0.07	0.03	*0.08*	-0.04
Agreeableness	IRT	-0.05	0.04	**0.13**	-0.02	0.02	0.03	-0.02	-0.02	-0.02	0.00	0.04	0.01	0.03	0.00	-0.03	*-0.08*
	Ipsative	-0.03	0.05	**0.16**	-0.02	0.02	0.01	-0.02	-0.01	-0.01	0.02	0.07	0.02	0.05	0.03	-0.04	*-0.08*
Emotional Stability	IRT	0.01	0.02	0.03	0.04	0.04	0.05	0.02	0.02	0.04	0.02	0.04	-0.01	0.02	**0.09**	0.00	0.02
	Ipsative	-0.01	0.01	0.05	0.03	0.03	0.06	0.03	0.01	0.03	0.03	0.04	-0.01	0.04	**0.10**	-0.04	0.01
Conscientiousness	IRT	0.03	0.05	*-0.08*	-0.04	0.02	0.06	-0.02	0.00	-0.03	-0.01	**0.13**	**0.10**	-0.05	-0.03	**0.13**	0.03
	Ipsative	-0.01	0.02	*-0.07*	*-0.09*	-0.03	0.06	-0.03	0.00	*-0.08*	-0.03	**0.15**	**0.12**	-0.04	-0.02	**0.10**	0.00

Bold – correlation significant at the 0.01 level (two-tailed)
Bold Italic – correlation significant at the 0.05 level (two-tailed)

OPQ32r criterion validity evidence

Study 1: OPQ32r as a predictor of supervisor performance in a retail setting (N=258)

This validation study was conducted within a large US-based retail organisation in order to identify personality constructs which could predict performance in supply chain supervisors.

Sample. Two hundred and fifty-eight supply chain supervisors from the company participated in the validation study. They were all located in the United States of America. Of the sample, 43.4% were female and 56.6% were male; 68.6% of the sample were White, 15.9% were Black/African American, 11.2% were Hispanic/Latino, and 1.9% were Asian, the remaining 2.4% were either American Indian/Alaskan Native, Native Hawaiian/Other Pacific Islander, or of mixed origin. Age ranged from 23 to 66, with an average sample age of 44.86. The median job tenure was 4 years (mean=3.34 years) and the median company tenure was 7.7 years (mean=13.36 years).

Instruments. A concurrent validity study was conducted using the OPQ32r as predictor of performance. The client organisation had developed a line-manager rating instrument consisting of 70 items intended to assess 12 areas of performance: Strategic Thinking, Business Knowledge, Decision Making, Innovation, Building Relationships, Developing Talent, Leading People, Communication, Customer Focus, Organising work, drive for results, and adaptability.

Method. The validation methodology described in chapter 8 (Criterion-related Validity) of the OPQ32 Technical Manual (SHL, 2006) was applied to the data obtained from OPQ32r. This is a theory-driven approach, which is commonly described as "criterion-centric" (Bartram, 2005).

Performance ratings by managers. As the behavioral statements rated by the line managers were not a part of a psychometrically validated tool, an investigation was carried out to identify constructs underlying the ratings. Both full information item factor analysis and principal components analysis of the 12 competency scales indicated the presence of just two factors, accounting for 73.4% of variance in the items and 90.5% in the scored competencies. The competencies loading on the first factor were: Strategic Thinking, Business Knowledge, Decision Making, Innovation, Organising Work, Drive for Results, and Adaptability. Those loading on the second factor were: Building Relationships, Developing Talent, Leading People, Communication, and Customer Focus. From the content it is clear that the first factor focuses on **Task Performance** while the second is focused on **Relationships with People**. In order to refine the criterion measure, items that had cross loadings on the two factors were removed and two shorter scales were produced. The first task-focused measure had 11 items and had an internal consistency (alpha) of 0.97. The second people-focused measure had nine items and an internal consistency (alpha) of 0.96. These two measures correlated 0.71 with each other.

The following section presents analyses of both the overall performance measure (i.e. the average of the 70 rating items) and on the two factor-based measures. Given the very high correlations between the 12 competency ratings, it is not meaningful to report on them individually.

Correlations between OPQ32r scale scores and criterion measure. Correlations were calculated between OPQ32r scale scores and the overall performance criterion measure. OPQ32r theta scores were used in all the analyses reported here due to the greater level of accuracy obtained through the use of thetas as opposed to derived standard scores (such as stens). In addition, individual scores on six broader factors, found to explain 57% of variance in OPQ32r scales, were calculated. Consistent with other datasets described in this manual, these factors were Extraversion, Agreeableness, Conscientiousness, Unconventionality, Critical Thinking and Emotional Stability (see in this manual for illustration).

A set of OPQ32r scales was hypothesised to relate to the overall performance ratings, on the basis of construct overlap between the competencies and the OPQ32r dimensions. These scales were: Data Rational, Evaluative, Caring, Outspoken, Achieving, Conscientious, and Persuasive. Correlations for non-hypothesised relationships were also calculated.

In addition, Great Eight and UCF 20 dimension competency potential scores were computed from the 32 OPQ32r scale scores using the standard equations. These competency potential scores were then correlated with the overall performance criterion and the criterion factor-based task and people scale scores.

Corrections for range restriction in the predictor were carried out. However this had relatively little effect on the uncorrected correlations. Typically it would increase a correlation of 0.18 to 0.20 and one of 0.27 to 0.30. For the present report, only uncorrected values are reported. However, corrected values would be approximately 10% larger.

Results. Table 24 shows the correlations between the OPQ32r theta scale scores and the overall performance rating scores, as well as with the Task Focus and People Focus criterion domain scores. This table also gives correlations between criteria and six broad factor scores derived from OPQ32r. It can be seen that there is very clear conceptual alignment between Task Focus and OPQ scales related to Task Performance (such as Evaluative, Data Rational, Detail Conscious and Achieving). Similarly, the People Focus competency is predicted by OPQ scales related to Relationships with People (such as Democratic and Persuasive). At the broad factor-level, the same strong conceptual alignment can be seen between predictor and performance; Task Focus is predicted by Critical Thinking (r=.27, p<.001) and People Focus by Agreeableness (r=.15, p<.05).

Results for the overall performance were in line with expectations: Data Rational predicted overall performance with r=0 .17 (p<.01) and Evaluative predicted overall performance with r=0.18 (p<.01). Outspoken predicted performance with r=0.12 (p<.05) and Achieving with r=0.15 (p<.05). Conscientiousness was not found to predict performance at the .05 level of significance, but did predict at the .10 level of significance (r=0.12, p=.055). Similarly, Persuasive (r=0.11, p=.09) did not predict at the .05 level of significance but did predict performance at the .10 level of significance. Finally, Caring (r=-.09, p=.17) was not found to be a predictor of overall performance. At the broader six-factor level, OPQ32r-derived scores for Critical Thinking (r=.18, p=0.005) and Conscientiousness (r=.11, p=0.08) were predictive of overall performance.

For non-hypothesised correlations, there were several significant relationships. First, Controlling was related to overall performance with r=0.15 (p<.05). Conceptual was also related to overall performance with r=0.14 (p<.05). Variety Seeking and Detail Consciousness were both related to overall performance with r=0.14 (p<.05). Finally, Conventional (r=-.11, p=.07) was a predictor of performance at the .10 level of significance.

Further analyses of the 20 competency potential predictor scores were carried out. Table 25 shows that the 20 UCF OPQ32r-based predictors clearly discriminate between those that predict Task Performance (8 competencies), those that predict People-Relationships (5 competencies) and those that predict neither (the remaining 7 competencies). Using stepwise regression, the OPQ32r scales Evaluative, Variety Seeking and Detail Conscious are related to the overall performance rating by R=0.28 (allowing for range restriction). When differentiated to pick up the task versus people focus aspects of the criterion, this increased to over R=0.35 (allowing for range restriction) for Task Performance and is around R=0.20 for people-related aspects of the competencies.

Table 24: Correlations between OPQ32r scores and performance ratings in a sample of supply chain managers (N=258).

OPQ32r scale	Task focus	People focus	Overall performance rating
Persuasive	.055	**.141**	**.130**
Controlling	**.134**	**.128**	**.174**
Outspoken	**.146**	.036	**.148**
Independent Minded	.046	**-.145**	-.012
Outgoing	-.098	.071	-.011
Affiliative	-.070	.026	-.037
Socially Confident	-.074	.090	.028
Modest	-.008	-.010	-.011
Democratic	.017	**.194**	.081
Caring	**-.151**	.058	-.088
Data Rational	**.256**	.012	**.173**
Evaluative	**.252**	.039	**.219**
Behavioral	-.018	.074	.025
Conventional	-.125	-.089	**-.145**
Conceptual	**.171**	.095	**.187**
Innovative	.064	.029	.090
Variety Seeking	**.147**	**.125**	**.168**
Adaptability	-.005	.012	.005
Forward Thinking	.034	.047	.054
Detail Conscious	**.167**	.057	**.136**
Conscientious	.111	.053	.116
Rule Following	-.025	-.005	-.033
Relaxed	.084	.028	.066
Worrying	-.056	-.039	-.053
Tough Minded	.117	.076	.097
Optimistic	-.075	.056	-.029
Trusting	-.048	.084	-.009
Emotionally Controlled	-.021	-.059	-.049
Vigorous	.050	.030	.051
Competitive	.060	.008	.069
Achieving	**.126**	.116	**.184**
Decisive	.049	-.045	.004
Broad factors:			
Extraversion	.052	.096	.114
Agreeableness	-.074	**.154**	.014
Emotional stability	.025	.088	.064
Unconventionality	.089	.065	.119
Critical thinking	**.268**	.049	**.219**
Conscientious	.101	.044	.096

Bold – correlations significant at the p=0.05 level

Table 25: UCF 20 competency potential scores (OPQ32r) correlated with overall criterion performance rating and with the separate factor-based Task and People focused measures (N=258).

	UCF 20 dimensions based on OPQ32r	Task focus	People focus	Overall performance rating
1.1	Deciding and Initiating Action	.09	.02	.09
1.2	Leading and Supervising	.00	**.19**	.09
2.1	Working with People	-.08	.12	-.01
2.2	Adhering to Principles and Values	-.09	.10	-.04
3.1	Relating and Networking	-.03	**.14**	.07
3.2	Persuading and Influencing	-.03	**.14**	.07
3.3	Presenting and Communicating Information	.11	**.13**	**.17**
4.1	Writing and Reporting	**.24**	.11	**.24**
4.2	Applying Expertise and Technology	**.26**	.05	**.23**
4.3	Analyzing	**.27**	.05	**.23**
5.1	Learning and Researching	**.29**	.08	**.25**
5.2	Creating and Innovating	**.14**	.06	**.16**
5.3	Formulating strategies and Concepts	.10	.08	**.13**
6.1	Planning and Organising	**.13**	.09	**.15**
6.2	Delivering Results and Meeting Customer Expectations	**.16**	.09	**.16**
6.3	Following Instructions and Procedures	.03	.06	.04
7.1	Adapting and Responding to Change	.09	**.13**	**.13**
7.2	Coping with Pressure and Setbacks	.05	.06	.04
8.1	Achieving personal Work Goals and Objectives	.12	.09	**.17**
8.2	Entrepreneurial and Commercial Thinking	**.19**	.07	**.20**

Bold – correlations significant at the p=0.05 level

Discussion. In summary, the OPQ32r shows a clear pattern of relationships with the criterion measures. The criterion is best considered as relating to two main aspects of performance: Task Performance and People Relationships. Considered in this way, the OPQ32r-based and 20 UCF competency-based measures of potential show clear discriminant validity, with different competency potential scores relating to each of the two criterion measures.

Study 2: OPQ32r as a predictor of Salon employee (hairdresser) performance in a retail setting (N=253)

This validation study was conducted within the same client organisation in Study 1: a large US-based retail organisation. The aim was to identify personality constructs which could predict performance for Salon employees (hairdressers).

Sample. Two hundred and fifty-three Salon employees from the company participated in the validation study. They were all located in the United States of America. Of the sample, 96.8% were female and 3.2% were male; 75.1% of the sample were White, 13.8% were Black/African American, 4.7% were Hispanic/Latino, and 2.0% were American Indian/Alaskan Native, the remaining 4.6% were either Asian, Native Hawaiian/Other Pacific Islander, or of mixed origin. Ages ranged from 22 to 69, with an average sample age of 43.92. The median job tenure was 1.83 years (mean=4 years) and the median company tenure was 4.51 years (mean=7.98 years).

Instruments. A concurrent validity study was conducted using the OPQ32r as a predictor of performance. The client organisation used the same criterion measurements as in Study 1, their own line-manager rating instrument consisting of 70 items intended to assess 12 areas of performance: Strategic Thinking, Business Knowledge, Decision Making, Innovation, Building Relationships, Developing Talent, Leading People, Communication, Customer Focus, Organising Work, Drive for Results, and Adaptability. For the present study, only scale scores were available, not item data.

Method. The same criterion-centric validation methodology as described earlier was followed.

Performance ratings by managers. As the behavioural statements rated by the line managers were not a part of a psychometrically validated tool, an investigation was carried out to identify constructs underlying the ratings. As the item data were not available, principal components analysis of the 12 competency scales was conducted. As in Study 1, this indicated the presence of just two factors, accounting for 62.45% of variance in the scored competencies. The competencies loading on the first

factor were: Strategic Thinking, Business Knowledge, Decision Making, Innovation, Organising Work, Drive for Results, and Adaptability. Those loading on the second factor were: Building Relationships, Developing Talent, Leading People, Communication, and Customer Focus. From the content, it is clear that the first factor focuses on **Task Performance** while the second is focused on **Relationships with People**. These two measures correlated 0.70 with each other.

The following report presents analyses of both the overall performance measure (i.e. the average of the 12 competency scales) and on the two factor-based measures. Given the very high correlations between the 12 competency ratings, it is not meaningful to report on them individually.

Correlations between OPQ32r scale scores and criterion measures. Correlations were calculated between OPQ32r scale scores and the overall performance criterion measure. A set of OPQ32r scales was hypothesised to relate to the overall performance ratings, on the basis of construct overlap between the competencies and the OPQ32r dimensions. These scales were: Data Rational, Evaluative, Caring, Outspoken, Achieving, Conscientious, and Persuasive. Correlations for non-hypothesised relationships were also calculated.

In addition, Great Eight and UCF 20 dimension competency potential scores were computed from the 32 OPQ32r scale scores using the standard equations. These competency potential scores were then correlated with the overall performance criterion and the criterion factor-based task and people scale scores.

Results and Discussion. Table 26 shows the correlations between the OPQ32r theta scale scores and the overall performance rating scores, as well as with the Task Focus and People Focus criterion domain scores. It can be seen that there is very clear conceptual alignment between Task Focus and OPQ scales related to Task Performance (such as Controlling, Modest, Data Rational, Detail Conscious and Conscientious). Similarly, People Focus competency is predicted by OPQ scales related to relationships with people (such as Persuasive and Behavioural). Optimistic and Achieving are significantly correlated with both criteria.

Results for the overall performance were in line with expectations: Data Rational predicted overall performance with r=0 .13 (p<.05). Conscientiousness predicted performance with r=.13 (p<.05). Finally, Achieving predicted performance with r=.18. For non-hypothesised relationships, Optimistic predicted performance with r=.20 and Worrying with r=-.13. Evaluative, Outspoken, Persuasive, and Caring were not significantly related to overall performance.

Stepwise multiple regression produced an R=0.26 for Achieving and Modest scales combined as significant predictors of the Task Focus criterion and R=0.31 for Achieving, Vigorous and Optimistic as significant predictors of the People Focus criterion. The overall aggregated criterion was predicted R=0.35 with Achieving, Vigorous, Optimistic, Conscientious and Modest as significant predictors.

Table 26: Correlations between OPQ32r scores and performance ratings in a sample of Salon employees (N=253).

OPQ32r scale	Task focus	People focus	Overall performance rating
Persuasive	.092	**.125**	.065
Controlling	**.121**	.079	.097
Outspoken	.021	-.033	.002
Independent Minded	.032	-.013	.020
Outgoing	.019	.050	.059
Affiliative	-.091	-.057	-.088
Socially Confident	.033	.048	.052
Modest	**.121**	-.034	.073
Democratic	.012	.031	.009
Caring	-.046	.076	-.026
Data Rational	**.132**	.071	**.132**
Evaluative	.111	.040	.082
Behavioral	.079	**.131**	.094
Conventional	-.056	-.012	-.014
Conceptual	.051	.096	.064
Innovative	.074	.082	.068
Variety Seeking	-.051	-.085	-.076
Adaptability	.003	.013	-.017
Forward Thinking	.111	.092	.097
Detail Conscious	**.131**	.042	.096
Conscientious	**.139**	.064	**.127**
Rule Following	-.043	-.002	-.018
Relaxed	.024	.066	.055
Worrying	-.116	-.100	-.131
Tough Minded	.036	-.008	-.026
Optimistic	**.145**	**.236**	**.200**
Trusting	-.045	.080	.017
Emotionally Controlled	-.027	-.065	-.054
Vigorous	-.053	-.081	-.115
Competitive	.088	.028	.045
Achieving	**.193**	**.198**	**.176**
Decisive	.091	-.016	.024

Bold – correlations significant at the p=0.05 level

Further analyses of the 20 competency potential predictor scores were carried out.

Table 27 shows that the 20 UCF OPQ32r based predictors discriminate between those that predict Task Performance (9 competencies), those that predict People-Relationships (1 competency) and those that predict neither (the remaining 10 competencies).

Table 27: UCF 20 competency potential scores (OPQ32r) correlated with overall criterion performance rating and with the separate factor based Task- and People-focused measures (N=253).

	UCF 20 dimensions based on OPQ32r	Task focus	People focus	Overall performance rating
1.1	Deciding and Initiating Action	.157	.079	.115
1.2	Leading and Supervising	.057	.156	.079
2.1	Working with People	-.025	.089	.008
2.2	Adhering to Principles and Values	-.069	.041	-.034
3.1	Relating and Networking	.060	.110	.079
3.2	Persuading and Influencing	.056	.119	.071
3.3	Presenting and Communicating Information	.111	.111	.110
4.1	Writing and Reporting	.143	.106	.115
4.2	Applying Expertise and Technology	.127	.087	.114
4.3	Analyzing	.126	.088	.117
5.1	Learning and Researching	.131	.081	.113
5.2	Creating and Innovating	.068	.047	.042
5.3	Formulating strategies and Concepts	.062	.095	.062
6.1	Planning and Organising	.174	.104	.145
6.2	Delivering Results and Meeting Customer Expectations	.143	.074	.116
6.3	Following Instructions and Procedures	.042	.035	.052
7.1	Adapting and Responding to Change	.051	.074	.039
7.2	Coping with Pressure and Setbacks	.077	.103	.078
8.1	Achieving personal Work Goals and Objectives	.139	.112	.090
8.2	Entrepreneurial and Commercial Thinking	.192	.117	.145

Bold – correlations significant at the p=0.05 level

Finally, analyses were conducted to examine the relationship between the Great 8 competencies and the Task Performance and People Relationships dimensions. Relationships between overall performance and the Great 8 were also examined. Table 28 outlines the findings of these analyses. Task Focus was significantly related to Leading and Deciding with r=.15 and Analysing and Interpreting with r=.13. People Focus was relating to Adapting and Coping with r=.13. Both Task Focus and People Focus were related to Enterprising and Performing with r=.18 and r=.16, respectively. Finally, Overall Performance was found to related to Leading and Deciding (r=.13), Analysing and Interpreting (r=.13), and Enterprising and Performing (r=.15).

Table 28: The Great 8 dimensions based on OPQ32r correlated with overall performance rating and with the separate factor based 'Task' and 'People' focused measures (N=253).

Great 8 dimensions based on OPQ32r	Task focus	People focus	Overall performance rating
Leading and Deciding	**.149**	.091	**.129**
Supporting and Cooperating	-.007	.105	.016
Interacting and Presenting	.057	.090	.063
Analyzing and Interpreting	**.126**	.096	**.125**
Creating and Conceptualising	.110	.108	.094
Organising and Executing	.086	.028	.064
Adapting and Coping	.081	**.133**	.093
Enterprising and Performing	**.182**	**.159**	**.150**

Bold – correlations significant at the p=0.05 level

Discussion. In summary, the OPQ32r shows a clear pattern of relationships with the criterion measures. The criterion is best considered as relating to two main aspects of performance: Task Performance and People Relationships. Considered in this way, the OPQ32r-based and 20 UCF competency-based measures of potential show discriminant validity, with different competency potential scores relating to each of the two criterion measures.

Stepwise multiple regression produced an R=0.26 for Achieving and Modest scales combined as significant predictors of the Task Focus criterion and R=0.31 for Achieving, Vigorous and Optimistic as significant predictors of the People Focus criterion. The overall aggregated criterion was predicted R=0.35 with Achieving, Vigorous, Optimistic, Conscientious and Modest as significant predictors.

General Discussion

The two studies confirm results presented in the OPQ32 Technical Manual (2006). Those studies were based on international data (including US data) and showed uncorrected correlations for hypothesised relationships between OPQ32 composite scale predictors and line manager competency ratings of around 0.20–0.25, reaching up to 0.45 for some competencies. Average correlations between non-hypothesised pairs of predictors and criteria were found to be around zero.

The present two validation studies were limited by the nature of the criterion measures. Although 70 items were rated by the managers, there was evidence of a very large halo effect, which resulted in it being impossible to resolve the original 12 competencies as separate measures. The data did afford the extraction of two scales, but even these were highly correlated (r=0.71 in Study 1 and r=0.70 in Study 2). Future studies need to employ forced-choice formats for criterion measures as well as for predictors (e.g. OPQ32r) in order to improve the possibility of demonstrating discriminant validity for manager-rated competencies (see Bartram, 2007).

Study 3: OPQ as a predictor of Front Line Supervisor performance (N=128)

A major manufacturer in North America wished to add formal selection assessments for front line supervisor positions. The organisation wanted to develop a selection tool that would assist them in selecting better candidates for this role. Specifically, the organisation wanted an online selection tool that would be cost-effective, job related, legally compliant, and would accurately identify candidates with the greatest performance potential.

Sample. The organisation identified incumbents for inclusion in the validation study. Responses to the tests were obtained from 135 incumbents and performance ratings were provided on 145 incumbents. Only cases representing matched (test and rating) data were retained for analyses. Based on this criterion, the final validation sample consisted of 128 incumbents.

Instruments. A concurrent validity study was conducted using the OPQ32r as a predictor of performance. The client organisation used their own line-manager rating instrument consisting of eight competencies intended to assess: Analysing and Problem Solving, Efficiency, Taking the Lead, Gaining Support, Inspiring Commitment, Development and Selection of Others, Communication Skills, Building Relationships, Having Credibility, and Being Adaptable. In order to predict these competencies, the OPQ scales were mapped onto the existing competency framework. Table 29 shows the mapping used to create the composites. For the purposes of this study, Analysing and Problem Solving and Development and Selection of Others were not included in the mapping (by request of the client).

Table 29: Competency and OPQ Dimension Mappings.

Competency	OPQ Dimensions
Execute Efficiently	Controlling, Decisive, Evaluative, Independent Minded, Detail Conscious
Show Initiative	Achieving, Decisive, Independent Minded, Outspoken
Solicit Support	Persuasive, Behavioural, Outgoing, Optimistic, Data Rational
Encourage Commitment	Emotional Control (-), Persuasive, Behavioural, Caring, Outgoing, Optimistic
Communicate Effectively	Behavioural, Persuasive, Evaluative, Democratic, Outspoken, Detail Oriented
Relates Well to Others	Relaxed, Socially Confident, Outgoing, Persuasive, Emotional Control
Demonstrates Credibility	Conscientious, Rule Following, Emotional Control, Democratic
Readily Adapt	Innovative, Conventional (-), Optimistic, Adaptable, Behavioural

Method. The same validation methodology as described earlier was applied in this study.

Performance Ratings by Managers. The performance ratings were completed by managers using an online tool which assessed all eight competencies. Specifically, managers provided global performance ratings for each employee under their supervision.

Results and Discussion. Table 30 shows the correlation between the predictor composite variables and the managerial competency ratings.

Table 30: Correlations between composite predictors and outcome variables.

Competency Dimension	Correlation with Composite OPQ Predictor
Execute Efficiently	.248**
Show Initiative	.231**
Solicit Support	.162
Encourage Commitment	.207*
Communicate Effectively	.187*
Relates Well to Others	.182*
Demonstrates Credibility	.132
Readily Adapt	.102

**p<.01, *p<.05.

An overall performance score was calculated as well, which was the sum of all performance dimensions. Table 31 demonstrates the relationships between the predictor composites and overall performance as rated by line managers.

Table 31: Correlations between composite predictors and overall performance.

Competency Dimension	Correlation with Overall Performance
Execute Efficiently	.251**
Show Initiative	.262**
Solicit Support	.213*
Encourage Commitment	.179
Communicate Effectively	.207*
Relates Well to Others	.160
Demonstrates Credibility	.220*
Readily Adapt	.277**

*p<.05, **p<.01

An overall predictor score was also calculated, which was the average of the OPQ predictor composites. This overall OPQ predictor composite correlated with overall performance at .260 (p<.01). This is an indicator of the highly predictive nature of the OPQ for overall performance for Front Line supervisors.

Discussion. Overall, the OPQ was found to significantly predict a majority of hypothesised competencies as well as overall performance for the role of Front Line Supervisors at a major manufacturer in North America. Further, overall performance was predicted by the overall composite variable, suggesting that the OPQ scales are powerful predictors of performance when used in combination with one another. Overall, this study demonstrates that the OPQ is a valid predictor of performance for Front Line Supervisor roles.

In summary, the IRT-scored OPQ32r is a valid instrument in predicting workplace criteria of various kinds. Generally, criterion-validity of its predecessor OPQ32i can be generalised to OPQ32r because relationships between the 32 scales and external criterion have been shown to be very similar for the two instruments. OPQ32r shows improvement in validity coefficients at the Big Five personality factors level because of unconstrained variance in composite scales derived from several convergent OPQ scales.

Chapter 6: Group Comparisons

Group comparisons on OPQ32r scale scores were conducted for Gender, Gender by Region, Education, Managerial Responsibility (manager and non-manager) and Managerial Responsibility by Region, using the international norm sample (sample 6) described in section 2.6. To avoid over- or underrepresentation by country or language, different countries or regions were weighted equally.

The comparisons followed effect size conventions suggested by Cohen (1988): small, 0.2 (0.4 sten), medium, 0.5 (1 sten) and large, 0.8 (1.6 sten). For practical applications, a medium effect is defined here as equal to or larger than 0.5 sten (d=0.25) and a large effect as equal to or exceeding 1.5 sten (d=0.75). Rounding these stens would result in 1 sten for a medium effect and 2 stens for a large effect. Effects smaller than 0.5 sten are considered as having little practical impact. Relationships with age correlations were carried out using age bands, and mean differences examined by converting age bands into two groups: over 40 and under 40. The information available for ethnicity varied greatly and was not comparable across the regions. For this reason information on ethnicity is not included in the group comparisons.

6.1 Gender differences

Gender differences were found on a number of scales, but the magnitude of these differences is typically small (below one sten). The largest differences were found for Competitive (effect size of 0.47), Vigorous (0.40) and Caring (0.34), which is in line with previous findings (SHL, 2006). Please see Table 32.

Gender by region

Gender differences were compared across regions. Effects sizes are given in Table 3333, with a positive value indicating that women have a higher score compare to men. Scales where the largest differences were found are presented graphically in Figures 8 to 11.

Across scales and regions we see that the largest positive difference (women scoring higher than men) occurs on Detail Conscious within the Icelandic sample, whereas the largest negative difference (Men scoring higher than women) is observed on Competitive within the New Zealand sample. Within the UK and US samples the largest positive differences are found for Caring and the largest negative differences for Competitive.

The within-region effect sizes were compared by correlating them against the effect sizes found in the overall sample. The pattern of effect sizes correlated highly for most regions (median=0.93, mean= 0.86). A small number of regions showed low correlations with the overall sample. The correlations ranged from 0.26 to 0.98, with lower correlations being found for: Korea (0.26), Middle East (0.39), Greater China (0.58) and China (0.73).

Table 32: Gender differences (effect sizes).

	N	Female			Male			Pooled SD	Effect Size	Abs Effect Size
		Weigh-ted %	Mean	SD	Weigh-ted %	Mean	SD			
Persuasive	118,324	36%	5.24	1.98	64%	5.65	1.97	1.97	-0.21	0.21*
Controlling	118,324	36%	5.23	2.01	64%	5.65	1.95	1.97	-0.22	0.22*
Outspoken	118,324	36%	5.39	2.01	64%	5.56	1.96	1.98	-0.09	0.09
Independent Minded	118,324	36%	5.46	1.94	64%	5.52	2.00	1.98	-0.03	0.03
Outgoing	118,324	36%	5.74	2.00	64%	5.36	1.95	1.97	0.19	0.19
Affiliative	118,324	36%	5.78	1.95	64%	5.34	1.98	1.97	0.22	0.22*
Socially Confident	118,324	36%	5.53	1.98	64%	5.48	1.98	1.98	0.02	0.02
Modest	118,324	36%	5.51	1.95	64%	5.50	2.00	1.98	0.01	0.01
Democratic	118,324	36%	5.62	1.96	64%	5.43	1.99	1.98	0.09	0.09
Caring	118,324	36%	5.93	1.93	64%	5.26	1.96	1.95	0.34	0.34*
Data Rational	118,324	36%	5.21	2.07	64%	5.66	1.91	1.97	-0.23	0.23
Evaluative	118,324	36%	5.43	2.00	64%	5.54	1.97	1.98	-0.06	0.06
Behavioural	118,324	36%	5.84	1.98	64%	5.31	1.95	1.96	0.27	0.27*
Conventional	118,324	36%	5.48	1.92	64%	5.51	2.01	1.98	-0.02	0.02
Conceptual	118,324	36%	5.46	1.96	64%	5.52	1.99	1.98	-0.03	0.03
Innovative	118,324	36%	5.30	2.02	64%	5.61	1.95	1.97	-0.16	0.16
Variety Seeking	118,324	36%	5.55	2.00	64%	5.47	1.97	1.98	0.04	0.04
Adaptable	118,324	36%	5.55	1.97	64%	5.47	1.98	1.98	0.04	0.04
Forward Thinking	118,324	36%	5.37	1.95	64%	5.57	1.99	1.98	-0.10	0.10
Detail Conscious	118,324	36%	5.85	1.98	64%	5.30	1.95	1.96	0.28	0.28*
Conscientious	118,324	36%	5.84	1.95	64%	5.31	1.97	1.96	0.27	0.27*
Rule Following	118,324	36%	5.51	1.91	64%	5.49	2.02	1.98	0.01	0.01
Relaxed	118,324	36%	5.32	1.95	64%	5.60	1.99	1.98	-0.15	0.15
Worrying	118,324	36%	5.86	1.94	64%	5.29	1.97	1.96	0.29	0.29
Tough Minded	118,324	36%	5.14	1.95	64%	5.70	1.97	1.96	-0.28	0.28
Optimistic	118,324	36%	5.67	1.99	64%	5.40	1.97	1.98	0.14	0.14
Trusting	118,324	36%	5.56	1.97	64%	5.46	1.99	1.98	0.05	0.05
Emotionally Controlled	118,324	36%	5.22	1.99	64%	5.66	1.96	1.97	-0.22	0.22
Vigorous	118,324	36%	6.00	1.93	64%	5.22	1.95	1.94	0.40	0.40
Competitive	118,324	36%	4.92	1.97	64%	5.83	1.91	1.93	-0.47	0.47
Achieving	118,324	36%	5.45	2.02	64%	5.53	1.96	1.98	-0.04	0.04
Decisive	118,324	36%	5.41	1.97	64%	5.55	1.98	1.98	-0.07	0.07

*Small effect size (0.2-0.5)

Table 33: Gender differences by region (effect sizes).

Absolute effect sizes are considered as below:

Small: 0.2-0.5	Medium: 0.5-0.8	Large: >0.8

	Australia	Belgium (Dutch)	Belgium (French)	Brazil	Canada	China	Denmark	Finland	France	Germany, Austria and Switzerland
N Male	5,488	1,400	1,568	732	423	2,028	4,440	2,791	2,636	1,397
N Female	3,632	985	961	274	280	1,294	2,369	2,590	1,589	939
Persuasive	-0.21	-0.13	-0.01	-0.53	-0.19	0.02	-0.34	-0.14	-0.23	-0.32
Controlling	-0.26	-0.33	-0.23	-0.19	-0.27	-0.04	-0.35	-0.32	-0.33	-0.37
Outspoken	-0.09	-0.15	-0.01	-0.02	-0.22	0.09	-0.11	-0.16	-0.11	-0.07
Independent Minded	-0.07	-0.06	0.10	0.15	-0.07	-0.06	-0.05	-0.02	-0.02	0.14
Outgoing	0.27	0.34	0.36	0.21	0.11	0.26	0.15	0.17	0.22	0.13
Affiliative	0.32	0.43	0.39	0.11	0.22	0.06	0.31	0.32	0.30	0.23
Socially Confident	0.06	0.04	0.10	-0.09	-0.13	0.24	-0.04	-0.07	-0.05	0.04
Modest	-0.08	-0.02	0.13	0.10	0.06	-0.05	0.10	-0.05	0.20	0.10
Democratic	0.14	0.28	0.07	-0.12	0.33	0.04	0.07	-0.02	0.03	0.04
Caring	0.49	0.58	0.37	0.19	0.47	0.21	0.39	0.42	0.29	0.31
Data Rational	-0.34	-0.28	-0.31	-0.52	-0.34	-0.02	-0.23	-0.09	-0.17	-0.28
Evaluative	-0.14	-0.23	-0.04	-0.03	-0.22	-0.02	-0.17	-0.21	-0.12	-0.05
Behavioural	0.30	0.46	0.33	0.18	0.39	0.18	0.15	0.33	0.13	0.30
Conventional	-0.04	0.06	-0.18	0.01	-0.06	-0.08	0.05	0.10	-0.01	-0.05
Conceptual	0.00	-0.16	-0.22	-0.06	-0.05	-0.07	0.00	-0.11	-0.08	-0.01
Innovative	-0.10	-0.23	-0.30	-0.29	-0.16	-0.08	-0.21	-0.28	-0.23	-0.23
Variety Seeking	0.01	-0.03	0.09	0.05	0.03	-0.06	-0.10	-0.10	0.06	-0.04
Adaptable	0.12	0.08	0.12	0.14	0.15	0.22	0.09	0.11	-0.02	0.15
Forward Thinking	0.00	0.00	-0.04	-0.44	-0.22	-0.04	-0.16	-0.23	-0.14	-0.24
Detail Conscious	0.44	0.40	0.41	-0.11	0.05	0.10	0.32	0.34	0.42	0.40
Conscientious	0.31	0.41	0.48	0.10	0.03	0.13	0.31	0.31	0.50	0.26
Rule Following	0.06	0.23	0.00	-0.23	-0.15	-0.08	0.18	0.21	0.20	0.20
Relaxed	-0.22	-0.36	-0.33	-0.06	-0.45	0.06	-0.18	-0.20	-0.26	-0.15
Worrying	0.31	0.33	0.29	0.29	0.61	0.07	0.33	0.30	0.35	0.44
Tough Minded	-0.39	-0.58	-0.32	-0.25	-0.49	-0.08	-0.33	-0.46	-0.26	-0.20
Optimistic	0.22	0.22	0.17	0.02	0.16	0.26	0.13	0.28	0.18	-0.03
Trusting	0.07	0.07	0.03	0.01	0.16	0.08	0.13	0.17	-0.08	-0.02
Emotionally Controlled	-0.26	-0.41	-0.28	-0.24	-0.20	-0.17	-0.24	-0.19	-0.13	-0.19
Vigorous	0.53	0.51	0.56	0.47	0.19	0.24	0.30	0.55	0.49	0.47
Competitive	-0.50	-0.39	-0.45	-0.52	-0.61	-0.25	-0.57	-0.51	-0.51	-0.52
Achieving	0.02	0.03	0.02	-0.24	-0.28	0.00	-0.18	-0.09	-0.09	-0.05
Decisive	-0.09	-0.24	-0.10	0.09	-0.05	0.01	-0.10	-0.09	-0.22	-0.10

	Greater China	Iceland	India	Italy	Korea	Latin America	Malaysia	Middle East	Netherlands	New Zealand
N Male	371	428	939	3,181	653	675	1,607	1,329	3,820	1,733
N Female	193	301	159	1,369	193	296	1,065	165	2,407	1,085
Persuasive	-0.25	-0.18	-0.36	-0.17	-0.29	-0.19	-0.07	-0.06	-0.31	-0.26
Controlling	-0.19	0.01	-0.16	-0.32	-0.40	-0.19	-0.27	0.19	-0.25	-0.33
Outspoken	-0.06	-0.07	-0.13	-0.05	-0.08	-0.04	0.06	-0.12	-0.12	-0.20
Independent Minded	-0.05	0.06	0.01	0.09	0.18	-0.27	0.07	0.16	0.08	-0.09
Outgoing	0.03	0.14	0.09	0.20	0.06	0.17	0.15	0.05	0.14	0.21
Affiliative	0.09	0.51	0.08	0.25	-0.22	0.19	0.10	0.11	0.26	0.32
Socially Confident	0.13	0.16	-0.02	0.02	-0.23	0.02	-0.07	-0.07	-0.14	0.01
Modest	0.03	-0.14	0.05	0.10	0.16	0.15	-0.02	-0.01	0.08	-0.06
Democratic	0.07	0.05	0.02	0.07	-0.34	-0.04	0.01	0.13	0.27	0.19
Caring	0.12	0.36	0.34	0.32	-0.07	0.35	0.16	0.15	0.47	0.46
Data Rational	-0.31	-0.11	-0.39	-0.26	-0.31	-0.22	-0.17	-0.30	-0.40	-0.21
Evaluative	-0.22	0.07	-0.04	0.04	-0.25	-0.15	-0.24	0.08	-0.13	-0.11
Behavioural	-0.04	0.42	0.28	0.32	-0.02	0.13	0.17	0.18	0.40	0.29
Conventional	0.14	0.04	-0.03	-0.15	0.12	0.18	-0.15	-0.38	0.07	0.02
Conceptual	-0.40	0.10	0.05	0.04	-0.25	-0.17	-0.06	-0.03	-0.10	-0.05
Innovative	-0.21	-0.23	-0.13	-0.17	-0.26	-0.31	-0.12	0.15	-0.13	-0.12
Variety Seeking	-0.10	0.06	0.13	0.12	0.11	-0.08	0.10	0.59	0.05	-0.04
Adaptable	0.05	0.08	0.05	0.01	-0.03	-0.20	0.11	0.02	0.10	0.08
Forward Thinking	-0.31	0.15	-0.24	-0.07	-0.27	-0.09	-0.18	0.03	-0.12	0.00
Detail Conscious	-0.11	0.69	0.28	0.19	-0.29	0.43	0.07	-0.03	0.29	0.50
Conscientious	-0.10	0.59	0.24	0.34	-0.24	0.26	0.09	0.02	0.25	0.38
Rule Following	0.03	0.18	0.08	-0.04	-0.22	0.07	-0.19	-0.21	0.13	0.12
Relaxed	0.09	-0.14	-0.12	-0.22	0.14	-0.18	-0.09	-0.16	-0.37	-0.28
Worrying	0.13	0.25	0.18	0.26	0.32	0.30	0.25	0.21	0.43	0.36
Tough Minded	-0.19	-0.21	-0.21	-0.26	-0.11	-0.13	-0.32	-0.08	-0.52	-0.40
Optimistic	0.09	0.37	-0.02	0.07	0.00	0.09	0.10	0.10	0.06	0.23
Trusting	0.29	0.29	-0.14	0.00	-0.04	0.07	0.02	0.15	0.13	0.16
Emotionally Controlled	-0.10	-0.32	-0.06	-0.20	0.11	-0.08	-0.15	0.07	-0.17	-0.22
Vigorous	0.30	0.48	0.19	0.35	0.20	0.33	0.31	0.37	0.43	0.52
Competitive	-0.19	-0.49	-0.34	-0.40	-0.14	-0.52	-0.23	-0.05	-0.56	-0.63
Achieving	-0.18	0.21	-0.11	0.00	-0.37	-0.15	0.01	0.19	-0.10	-0.02
Decisive	-0.02	-0.09	-0.26	-0.16	0.09	-0.05	-0.07	-0.23	-0.16	-0.08

	Norway	Poland	Portugal	Singapore	South Africa	Spain	Sweden	Turkey	UK	US
N Male	3,050	594	1,264	2,301	2,545	485	7,744	623	13,801	2,588
N Female	1,609	333	984	1,752	2,335	210	5,500	484	8,811	1,526
Persuasive	-0.26	-0.36	-0.22	-0.38	-0.17	-0.29	-0.24	0.03	-0.16	-0.20
Controlling	-0.11	-0.35	-0.28	-0.33	-0.19	-0.29	-0.10	0.04	-0.22	-0.16
Outspoken	-0.08	-0.01	-0.16	-0.06	-0.01	-0.07	-0.08	0.07	-0.20	-0.08
Independent Minded	-0.12	-0.20	-0.05	-0.14	-0.04	0.00	0.03	-0.02	-0.20	-0.02
Outgoing	0.18	0.11	0.25	-0.06	0.18	0.29	0.16	0.44	0.23	0.12
Affiliative	0.30	0.41	0.15	0.11	0.09	0.15	0.15	0.34	0.30	0.15
Socially Confident	-0.05	-0.11	-0.03	-0.12	0.04	-0.05	0.00	0.07	0.03	-0.04
Modest	0.00	0.00	0.01	0.05	-0.04	0.04	0.02	-0.31	0.00	-0.08
Democratic	0.16	0.23	0.04	0.13	-0.11	-0.07	0.05	-0.01	0.14	-0.05
Caring	0.38	0.33	0.30	0.17	0.25	0.24	0.34	0.38	0.48	0.39
Data Rational	-0.15	-0.22	-0.27	-0.11	-0.23	-0.25	-0.17	-0.15	-0.29	-0.14
Evaluative	-0.11	-0.07	-0.13	-0.22	-0.26	-0.04	-0.09	0.04	-0.16	-0.12
Behavioural	0.28	0.36	0.19	0.15	0.10	0.15	0.28	0.31	0.32	0.25
Conventional	-0.03	-0.06	-0.13	0.05	0.08	0.05	0.03	-0.18	0.01	0.03
Conceptual	0.04	-0.05	-0.13	-0.11	-0.13	0.06	-0.01	-0.10	-0.08	-0.07
Innovative	-0.12	-0.20	-0.18	-0.30	-0.18	-0.13	-0.11	0.00	-0.15	-0.12
Variety Seeking	-0.05	0.08	0.05	-0.05	-0.04	0.02	-0.01	0.22	-0.04	-0.06
Adaptable	0.08	0.02	0.07	0.15	0.03	0.15	0.10	-0.04	0.12	0.11
Forward Thinking	-0.07	-0.21	-0.13	-0.16	-0.17	-0.02	-0.03	0.00	-0.06	0.07
Detail Conscious	0.31	0.22	0.34	0.26	0.29	0.26	0.39	0.19	0.37	0.35
Conscientious	0.33	0.11	0.26	0.19	0.22	0.25	0.32	0.35	0.28	0.25
Rule Following	0.12	0.07	0.06	0.01	0.12	0.00	0.20	-0.11	0.10	0.15
Relaxed	-0.14	-0.22	-0.19	-0.17	-0.14	-0.31	-0.24	-0.19	-0.29	-0.24
Worrying	0.33	0.42	0.47	0.36	0.26	0.49	0.30	0.37	0.41	0.38
Tough Minded	-0.38	-0.32	-0.40	-0.28	-0.16	-0.41	-0.41	-0.30	-0.37	-0.25
Optimistic	0.23	0.01	0.14	0.05	0.06	0.12	0.26	0.12	0.14	0.15
Trusting	0.18	0.09	-0.08	0.07	-0.10	0.03	0.17	-0.20	0.04	0.02
Emotionally Controlled	-0.24	-0.26	-0.18	-0.15	-0.18	-0.21	-0.27	-0.43	-0.22	-0.22
Vigorous	0.39	0.50	0.48	0.25	0.42	0.43	0.36	0.54	0.46	0.28
Competitive	-0.52	-0.53	-0.52	-0.44	-0.29	-0.57	-0.49	-0.15	-0.55	-0.40
Achieving	0.04	-0.05	-0.06	-0.24	0.03	-0.05	0.02	0.13	-0.01	0.01
Decisive	-0.07	-0.17	0.04	-0.14	-0.14	-0.03	0.02	-0.09	-0.12	-0.12

Figure 8: Gender differences on Detail Conscious (positive effect sizes indicate women score higher than men).

Detail Conscious
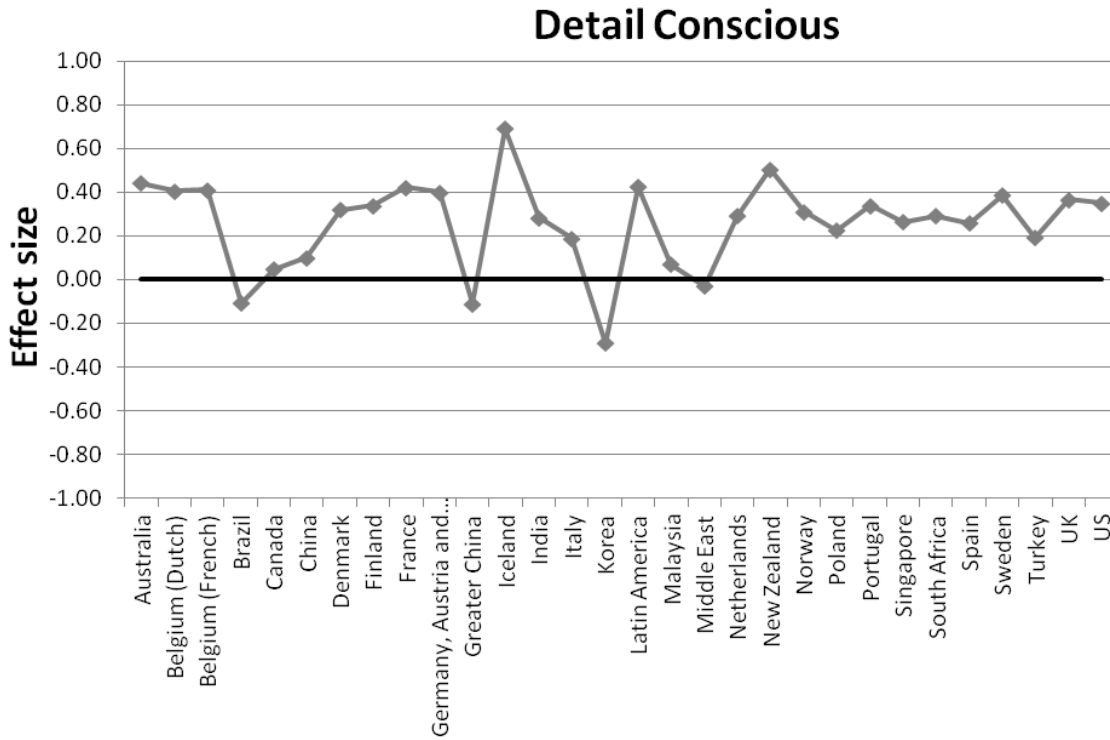
Figure 9: Gender differences on Conscientious (positive effect sizes indicate women score higher than men).

Conscientious
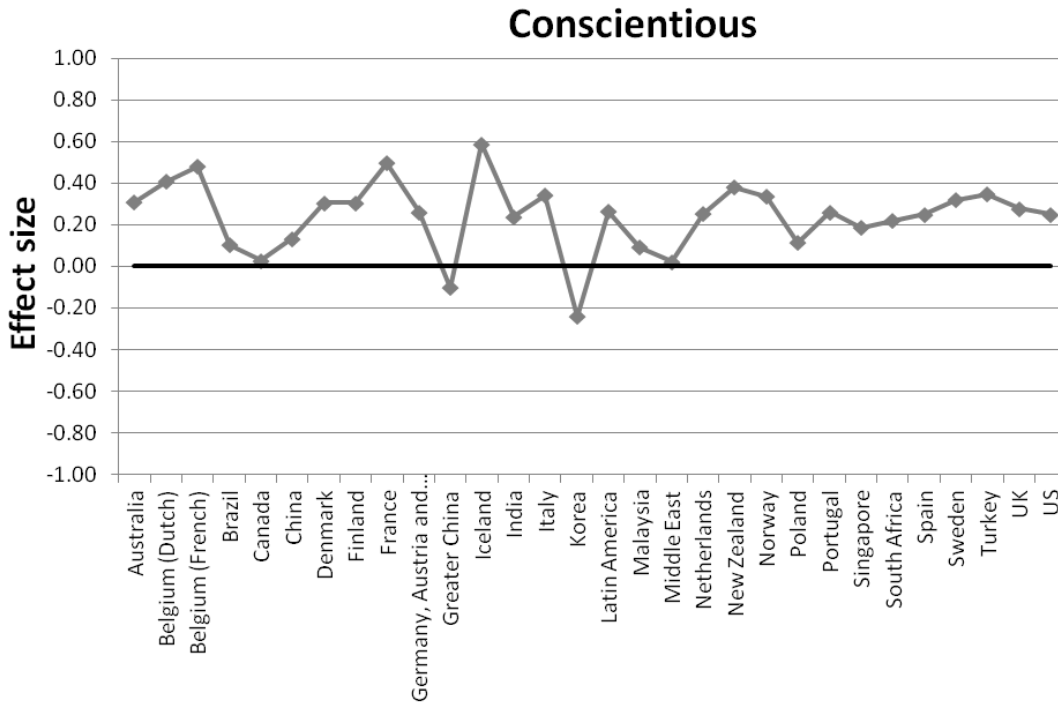
Figure 10: Gender differences on Competitive (positive effect sizes indicate women score higher than men).
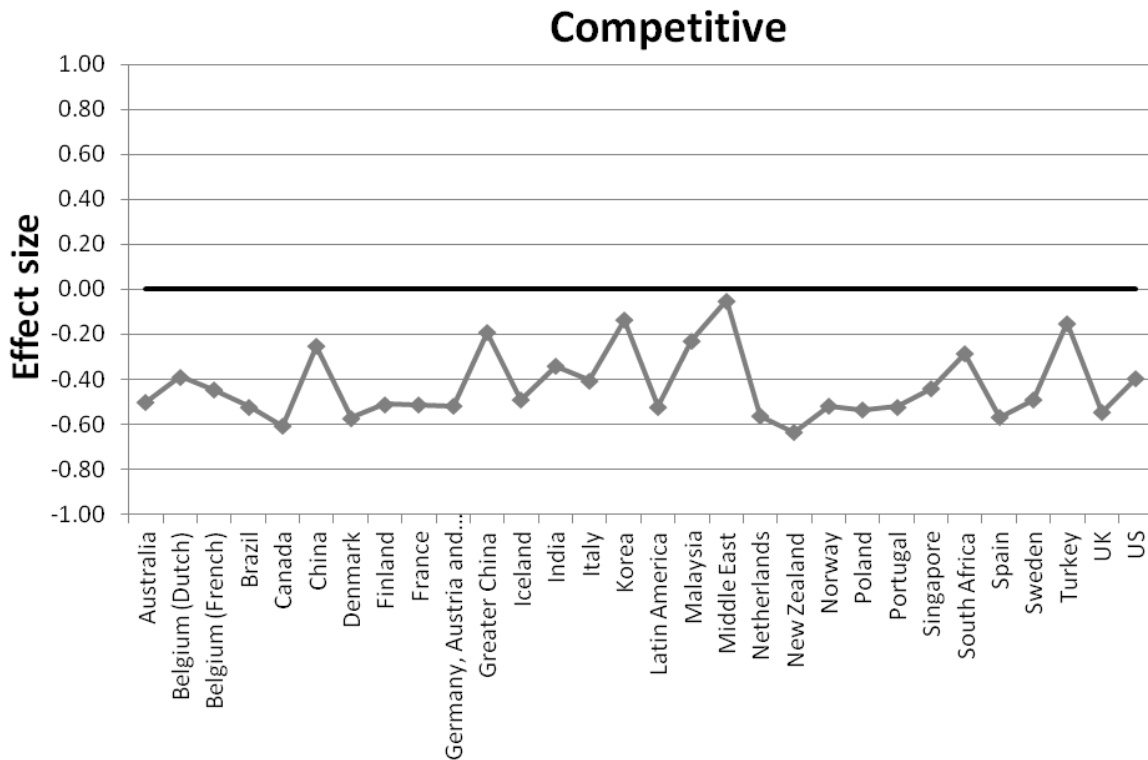
Figure 11: Gender differences on Affiliative (positive effect sizes indicate women score higher than men).
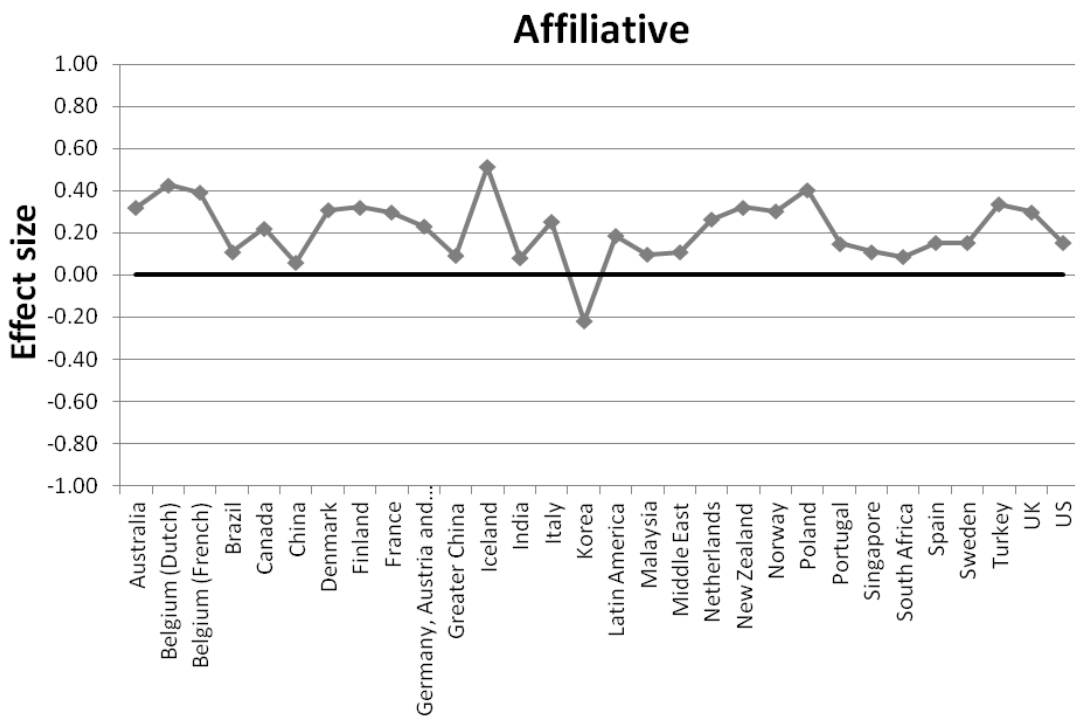
6.2 Age

Correlations between age and scale scores of the OPQ32r were small. Across all scales, the average absolute correlation was 0.03. The largest correlation with age was found for the Trusting scale (0.11) indicating older people rated themselves as more trusting compared to younger people. Correlations and descriptive statistics can be found in Table 34Table 34. All correlations above 0.03 are significant at $p < 0.001$ as a result of the large sample size.

Table 34: Correlations with age.

	N	Mean	SD	Correlation coefficient
Persuasive	79,149	5.50	1.98	0.04
Controlling	79,149	5.50	1.98	0.04
Outspoken	79,149	5.50	1.98	0.01
Independent Minded	79,149	5.50	1.98	0.04
Outgoing	79,149	5.50	1.98	-0.04
Affiliative	79,149	5.50	1.98	-0.03
Socially Confident	79,149	5.50	1.98	0.00
Modest	79,149	5.50	1.98	0.03
Democratic	79,149	5.50	1.98	0.02
Caring	79,149	5.50	1.98	0.03
Data Rational	79,149	5.50	1.98	-0.07
Evaluative	79,149	5.50	1.98	-0.02
Behavioural	79,149	5.50	1.98	0.00
Conventional	79,149	5.50	1.98	0.00
Conceptual	79,149	5.50	1.98	-0.06
Innovative	79,149	5.50	1.98	0.00
Variety Seeking	79,149	5.50	1.98	0.04
Adaptable	79,149	5.50	1.98	0.01
Forward Thinking	79,149	5.50	1.98	-0.01
Detail Conscious	79,149	5.50	1.98	-0.05
Conscientious	79,149	5.50	1.98	-0.04
Rule Following	79,149	5.50	1.98	-0.02
Relaxed	79,149	5.50	1.98	0.01
Worrying	79,149	5.50	1.98	-0.03
Tough Minded	79,149	5.50	1.98	-0.01
Optimistic	79,149	5.50	1.98	0.05
Trusting	79,149	5.50	1.98	0.11
Emotionally Controlled	79,149	5.50	1.98	0.02
Vigorous	79,149	5.50	1.98	-0.02
Competitive	79,149	5.50	1.98	-0.05
Achieving	79,149	5.50	1.98	-0.07
Decisive	79,149	5.50	1.98	0.10

Comparison for groups above and below age 40

Mean scale scores of individuals below 40 and 40 and above were also compared. A number of small effect sizes were found, indicating that people above 40 were more Trusting (effect size 0.39), more Decisive (0.27), less Achieving (0.23) and less Data Rational (0.22). Differences are typically small (below one sten). These findings are in line with previous findings on relationships with age. Further details can be found in Table 35.

Table 35: Mean differences and effect size on age (above and below 40).

	N	Below 40			Above 40			Pooled SD	Effect Size	Abs Effect Size
		Weigh-ted %	Mean	SD	Weigh-ted %	Mean	SD			
Persuasive	79,149	73%	5.49	1.99	27%	5.76	1.96	1.98	-0.136	0.14
Controlling	79,149	73%	5.46	2.00	27%	5.78	1.97	1.99	-0.162	0.16
Outspoken	79,149	73%	5.59	1.99	27%	5.60	1.93	1.98	-0.004	0.00
Independent Minded	79,149	73%	5.51	1.97	27%	5.67	1.96	1.97	-0.079	0.08
Outgoing	79,149	73%	5.61	1.98	27%	5.38	1.95	1.97	0.116	0.12
Affiliative	79,149	73%	5.61	1.97	27%	5.42	1.96	1.97	0.095	0.10
Socially Confident	79,149	73%	5.50	1.97	27%	5.58	1.97	1.97	-0.041	0.04
Modest	79,149	73%	5.37	1.98	27%	5.64	1.93	1.97	-0.136	0.14
Democratic	79,149	73%	5.44	1.96	27%	5.63	2.04	1.98	-0.092	0.09
Caring	79,149	73%	5.44	1.99	27%	5.65	1.96	1.98	-0.107	0.11
Data Rational	79,149	73%	5.64	2.02	27%	5.21	1.86	1.97	0.220	0.22*
Evaluative	79,149	73%	5.67	1.96	27%	5.53	1.91	1.95	0.070	0.07
Behavioural	79,149	73%	5.53	1.99	27%	5.51	1.94	1.98	0.011	0.01
Conventional	79,149	73%	5.38	1.93	27%	5.45	2.02	1.96	-0.036	0.04
Conceptual	79,149	73%	5.59	1.97	27%	5.22	2.02	1.98	0.187	0.19
Innovative	79,149	73%	5.55	1.99	27%	5.64	1.97	1.98	-0.047	0.05
Variety Seeking	79,149	73%	5.57	1.99	27%	5.74	1.86	1.96	-0.083	0.08
Adaptable	79,149	73%	5.48	2.01	27%	5.35	1.87	1.97	0.064	0.06
Forward Thinking	79,149	73%	5.47	1.99	27%	5.47	1.98	1.98	0.000	0.00
Detail Conscious	79,149	73%	5.59	1.98	27%	5.28	1.93	1.96	0.158	0.16
Conscientious	79,149	73%	5.63	1.99	27%	5.43	1.96	1.98	0.099	0.10
Rule Following	79,149	73%	5.56	1.93	27%	5.49	2.00	1.95	0.036	0.04
Relaxed	79,149	73%	5.53	1.98	27%	5.63	1.99	1.98	-0.049	0.05
Worrying	79,149	73%	5.55	2.00	27%	5.28	1.94	1.98	0.134	0.13
Tough Minded	79,149	73%	5.48	1.99	27%	5.53	1.93	1.98	-0.025	0.03
Optimistic	79,149	73%	5.41	1.99	27%	5.70	1.99	1.99	-0.148	0.15
Trusting	79,149	73%	5.29	1.95	27%	6.06	1.95	1.95	-0.392	0.39*
Emotionally Controlled	79,149	73%	5.44	2.00	27%	5.59	1.92	1.98	-0.073	0.07
Vigorous	79,149	73%	5.62	1.98	27%	5.55	1.87	1.96	0.036	0.04
Competitive	79,149	73%	5.57	1.99	27%	5.29	2.03	2.00	0.136	0.14
Achieving	79,149	73%	5.71	1.97	27%	5.24	1.97	1.97	0.235	0.24*
Decisive	79,149	73%	5.46	1.93	27%	5.99	1.95	1.94	-0.278	0.28*

*Small effect size (0.2-0.5)

6.3 Ethnic background (UK and US data)

UK data

Scores across ethnic groups were compared using the OPQ32r UKE General Population Norm Sample (N=22,612). Limited ethnicity data was available and for the purpose of this study ethnicity data was recoded into white and non-white. Table 36 shows effect sizes by OPQ scales. The magnitude of the differences is typically small (below one sten). The highest absolute difference observed is d=0.39 for the scale Rule Following, which suggests that the White ethnicity group tend to see themselves as more Rule Following than the Non-white group. This does not exceed a small effect size.

Table 36: Mean differences and effect size by ethnicity (White – Non-white).

OPQ32r scale	Non-white (N=12,602)		White (N=1,758)		Effect size
	Mean	SD	Mean	SD	
Persuasive	5.52	1.96	5.54	1.92	0.01
Controlling	5.63	1.99	5.29	1.80	-0.17
Outspoken	5.29	1.98	4.95	1.84	-0.17
Independent Minded	5.20	1.98	4.91	1.94	-0.14
Outgoing	5.41	2.09	5.17	1.81	-0.12
Affiliative	5.30	1.97	5.01	1.78	-0.15
Socially Confident	5.86	2.13	5.99	1.91	0.06
Modest	5.81	2.06	5.62	1.98	-0.09
Democratic	5.92	2.09	5.66	1.95	-0.13
Caring	5.73	2.10	5.76	1.99	0.01
Data Rational	5.44	2.11	6.19	2.11	0.35*
Evaluative	5.91	2.01	5.80	1.92	-0.05
Behavioural	5.79	2.07	5.32	1.88	-0.23*
Conventional	5.47	2.11	6.00	1.97	0.26*
Conceptual	5.78	2.13	5.88	1.85	0.04
Innovative	5.63	2.01	5.68	1.86	0.02
Variety Seeking	5.42	2.05	4.86	1.92	-0.28*
Adaptable	5.63	2.07	5.56	1.90	-0.04
Forward Thinking	5.34	1.99	5.81	1.84	0.24*
Detail Conscious	5.75	2.10	6.21	1.91	0.22*
Conscientious	5.73	2.07	5.94	1.92	0.10
Rule Following	5.05	2.07	5.85	2.04	0.39*
Relaxed	5.43	2.12	5.78	1.96	0.16
Worrying	5.55	2.08	5.19	1.92	-0.18
Tough Minded	5.71	2.16	5.77	1.99	0.03
Optimistic	5.23	1.99	5.42	1.84	0.10
Trusting	5.36	1.84	5.12	1.82	-0.13
Emotionally Controlled	5.67	2.01	5.73	1.81	0.03
Vigorous	6.32	1.76	6.05	1.79	-0.15
Competitive	5.20	2.17	5.43	1.94	0.11
Achieving	5.59	1.99	6.06	1.87	0.24*
Decisive	5.20	1.99	4.96	1.85	-0.12

*Small effect size (0.2-0.5)

US data

Scores across ethnicity groups were compared within a sample collected between December 2009 and July 2010, using the OPQ32r (N = 2,473). Of the respondents, 86% were from the US and 16% from Canada. Information on ethnic background was available for 43% (N = 1,060) and for the purpose of this study ethnicity data was recoded into white and minority combined.

Table 37 compares OPQ32r scores of the Minority Combined group with those of the White group. The magnitude of the differences is typically small (below one sten). The highest absolute difference observed is d=0.34 for the scale Affiliative, which suggests that the White ethnicity group tends to see themselves as more Affiliative than the Minority Combined group. This does not exceed a small effect size.

Table 37: Mean differences and effect size by ethnicity (Minority combined – White).

OPQ32r scale	Minority combined (N = 222)		White (N = 838)		Effect Size
	Mean	**SD**	**Mean**	**SD**	
Persuasive	4.77	1.77	4.92	1.94	-0.08
Controlling	4.88	1.78	5.20	1.97	-0.16
Outspoken	4.97	1.80	5.33	2.03	-0.18
Independent Minded	5.40	2.06	5.54	1.97	-0.07
Outgoing	5.19	1.79	5.35	2.09	-0.08
Affiliative	4.66	1.88	5.36	2.13	-0.34*
Socially Confident	4.92	1.88	5.08	2.02	-0.08
Modest	5.72	1.87	5.82	1.99	-0.05
Democratic	4.91	1.75	5.29	1.96	-0.2
Caring	5.44	1.77	5.29	1.91	0.08
Data Rational	5.65	2.00	5.30	2.05	0.18
Evaluative	5.08	1.90	5.10	2.08	-0.01
Behavioral	5.11	1.69	5.13	2.00	-0.01
Conventional	6.17	1.80	5.75	2.00	0.21
Conceptual	5.48	1.78	5.26	2.04	0.11
Innovative	5.15	1.86	5.18	2.03	-0.01
Variety Seeking	5.16	1.91	5.49	1.92	-0.17
Adaptable	5.36	1.77	5.41	1.99	-0.02
Forward Thinking	5.36	1.96	5.11	2.03	0.12
Detail Conscious	5.57	1.89	5.35	1.95	0.12
Conscientious	5.61	1.93	5.44	1.95	0.09
Rule Following	6.05	1.72	5.56	1.91	0.27*
Relaxed	5.94	2.02	5.27	2.06	0.33*
Worrying	5.88	1.86	5.81	2.01	0.04
Tough Minded	5.19	1.83	5.25	2.01	-0.03
Optimistic	5.36	1.74	5.19	2.07	0.09
Trusting	5.22	1.94	5.25	2.07	-0.02
Emotionally Controlled	5.92	1.80	5.78	2.04	0.07
Vigorous	5.49	1.94	5.68	1.95	-0.1
Competitive	5.09	1.56	5.37	1.92	-0.15
Achieving	5.22	1.80	5.03	2.02	0.1
Decisive	5.09	1.78	5.48	2.00	-0.2

*Small effect size (0.2-0.5)

6.4 Education: Above or below degree level

OPQ32r scores of individuals who obtained a university degree (or additional postgraduate degrees) were compared to those without a degree. A number of small effect sizes were found, indicating that individuals with a degree were more Evaluative (effect size 0.37), more Forward thinking (0.31), more Achieving (0.28) more Data Rational (0.24), more Variety Seeking (0.23), less Conventional (0.22) and more Controlling (0.21). Mean scores and effect size can be found in Table 38.

Table 38: Mean differences and effect sizes on education.

	N	Degree Level or Above			No Degree			Pooled SD	Effect Size	Abs Effect Size
		Weigh-ted %	Mean	SD	Weigh-ted %	Mean	SD			
Persuasive	52,099	70%	5.66	2.00	30%	5.48	1.99	2.00	0.090	0.09
Controlling	52,099	70%	5.67	1.96	30%	5.26	2.07	1.99	0.208	0.21*
Outspoken	52,099	70%	5.56	1.96	30%	5.54	2.04	1.99	0.009	0.01
Independent Minded	52,099	70%	5.56	1.96	30%	5.55	1.98	1.97	0.004	0.00
Outgoing	52,099	70%	5.53	1.98	30%	5.78	1.98	1.98	-0.128	0.13
Affiliative	52,099	70%	5.60	1.98	30%	5.60	1.96	1.97	0.003	0.00
Socially Confident	52,099	70%	5.60	1.97	30%	5.52	1.97	1.97	0.040	0.04
Modest	52,099	70%	5.43	1.98	30%	5.46	1.95	1.97	-0.017	0.02
Democratic	52,099	70%	5.57	2.01	30%	5.26	1.92	1.98	0.157	0.16
Caring	52,099	70%	5.49	1.99	30%	5.54	1.99	1.99	-0.028	0.03
Data Rational	52,099	70%	5.70	2.00	30%	5.23	1.94	1.98	0.239	0.24*
Evaluative	52,099	70%	5.82	1.92	30%	5.12	1.96	1.93	0.367	0.37*
Behavioural	52,099	70%	5.60	1.99	30%	5.40	1.94	1.97	0.103	0.10
Conventional	52,099	70%	5.29	1.96	30%	5.72	1.92	1.95	-0.220	0.22*
Conceptual	52,099	70%	5.57	1.98	30%	5.24	1.97	1.98	0.164	0.16
Innovative	52,099	70%	5.69	1.99	30%	5.31	1.96	1.98	0.187	0.19
Variety Seeking	52,099	70%	5.73	1.99	30%	5.29	1.90	1.96	0.225	0.23*
Adaptable	52,099	70%	5.46	1.99	30%	5.37	1.96	1.98	0.044	0.04
Forward Thinking	52,099	70%	5.71	1.98	30%	5.09	1.94	1.97	0.313	0.31*
Detail Conscious	52,099	70%	5.49	1.98	30%	5.45	1.95	1.97	0.019	0.02
Conscientious	52,099	70%	5.60	1.97	30%	5.47	2.01	1.98	0.069	0.07
Rule Following	52,099	70%	5.42	1.97	30%	5.73	1.91	1.95	-0.156	0.16
Relaxed	52,099	70%	5.49	1.98	30%	5.82	2.00	1.98	-0.166	0.17
Worrying	52,099	70%	5.37	1.97	30%	5.48	1.97	1.97	-0.053	0.05
Tough Minded	52,099	70%	5.55	1.99	30%	5.48	1.97	1.98	0.035	0.04
Optimistic	52,099	70%	5.57	2.01	30%	5.50	1.94	1.99	0.035	0.04
Trusting	52,099	70%	5.60	1.96	30%	5.38	2.05	1.99	0.110	0.11
Emotionally	52,099	70%	5.42	1.97	30%	5.57	2.00	1.98	-0.076	0.08
Vigorous	52,099	70%	5.64	1.95	30%	5.71	1.99	1.96	-0.036	0.04
Competitive	52,099	70%	5.53	2.01	30%	5.43	2.06	2.02	0.054	0.05
Achieving	52,099	70%	5.77	1.95	30%	5.22	2.00	1.97	0.278	0.28*
Decisive	52,099	70%	5.66	1.99	30%	5.58	1.90	1.96	0.040	0.04

*Small effect size (0.2-0.5)

6.5 Job Level: Managers and non-Managers

When comparing scores of managers with non-managers a number of small and medium effect sizes were found. These results indicated managers were more Controlling (effect size 0.56), more Decisive (0.40), less Affiliative (0.29), more Forward Thinking (0.25) and more Persuasive (0.22). Further details can be found in Table 39.

Table 39: Mean differences and effect sizes based on managerial role.

	N	Managers			Non-Managers			Pooled SD	Effect Size	Abs Effect Size
		Weigh-ted %	Mean	SD	Weigh-ted %	Mean	SD			
Persuasive	44,625	40%	5.87	1.93	60%	5.44	2.03	1.99	0.216	0.22*
Controlling	44,625	40%	6.24	1.79	60%	5.14	2.05	1.95	0.562	0.56**
Outspoken	44,625	40%	5.77	1.92	60%	5.42	2.04	2.00	0.175	0.17
Independent Minded	44,625	40%	5.75	1.95	60%	5.44	2.00	1.98	0.157	0.16
Outgoing	44,625	40%	5.46	1.96	60%	5.74	2.02	2.00	-0.139	0.14
Affiliative	44,625	40%	5.23	1.96	60%	5.80	1.98	1.97	-0.289	0.29*
Socially Confident	44,625	40%	5.64	1.94	60%	5.68	2.01	1.98	-0.019	0.02
Modest	44,625	40%	5.48	1.94	60%	5.36	2.01	1.98	0.058	0.06
Democratic	44,625	40%	5.53	2.04	60%	5.53	1.96	2.00	0.004	0.00
Caring	44,625	40%	5.45	1.97	60%	5.59	2.02	2.00	-0.066	0.07
Data Rational	44,625	40%	5.55	1.91	60%	5.54	2.07	2.01	0.005	0.00
Evaluative	44,625	40%	5.63	1.88	60%	5.49	2.02	1.97	0.071	0.07
Behavioural	44,625	40%	5.55	1.91	60%	5.51	2.01	1.97	0.017	0.02
Conventional	44,625	40%	5.28	1.96	60%	5.57	1.94	1.95	-0.146	0.15
Conceptual	44,625	40%	5.41	1.96	60%	5.57	2.00	1.98	-0.081	0.08
Innovative	44,625	40%	5.84	1.92	60%	5.46	2.03	1.99	0.187	0.19
Variety Seeking	44,625	40%	5.73	1.88	60%	5.43	2.01	1.96	0.156	0.16
Adaptable	44,625	40%	5.37	1.92	60%	5.38	2.02	1.98	-0.003	0.00
Forward Thinking	44,625	40%	5.89	1.96	60%	5.39	2.01	1.99	0.250	0.25*
Detail Conscious	44,625	40%	5.32	1.93	60%	5.61	1.98	1.96	-0.148	0.15
Conscientious	44,625	40%	5.38	1.96	60%	5.67	2.02	1.99	-0.148	0.15
Rule Following	44,625	40%	5.35	1.98	60%	5.61	1.97	1.97	-0.136	0.14
Relaxed	44,625	40%	5.43	1.97	60%	5.82	1.99	1.98	-0.197	0.20*
Worrying	44,625	40%	5.12	1.89	60%	5.36	1.99	1.95	-0.122	0.12
Tough Minded	44,625	40%	5.57	1.93	60%	5.59	2.01	1.98	-0.007	0.01
Optimistic	44,625	40%	5.75	1.95	60%	5.57	2.00	1.98	0.093	0.09
Trusting	44,625	40%	5.70	1.94	60%	5.47	2.03	1.99	0.118	0.12
Emotionally Controlled	4,462	40%	5.45	1.92	60%	5.39	2.02	1.98	0.028	0.03
Vigorous	44,625	40%	5.53	1.89	60%	5.83	2.01	1.96	-0.151	0.15
Competitive	44,625	40%	5.74	1.94	60%	5.39	2.09	2.03	0.170	0.17
Achieving	44,625	40%	5.83	1.91	60%	5.53	2.03	1.98	0.154	0.15
Decisive	44,625	40%	6.06	1.94	60%	5.30	1.92	1.93	0.396	0.40*

*Small effect size (0.2-0.5).
**Medium effect size (0.5-0.8)

Managers and Non-managers by Region

Scale scores of managers and non-managers were compared across regions, where sufficient information was available on managerial status. Effect sizes are given in Table 40, with a positive value indicating that managers have a higher score compared to non-managers. The scales with largest spread in mean personality scores are also displayed graphically in Figures 12 to 15. Across scales and regions we see that the largest positive difference (managers scoring higher than non-managers) occurs on Controlling within the Finnish sample. Whereas the largest negative difference (non-managers scoring higher than managers) is observed on Rule following within the UK sample.

The within-region effect sizes were compared by correlating them against the effect sizes found in the overall sample. The pattern of effect sizes correlated highly for most regions (median=0.83, mean=0.81). A small number of regions showed low correlations with the overall sample. The correlations ranged from 0.52 to 0.89, with lower correlations being found for: Middle East (0.52), Malaysia (0.65) and India (0.70).

Table 40: Managerial Role differences by region (effect sizes).

Absolute effect sizes are considered as below:

Small: 0.2-0.5	Medium: 0.5-0.8	Large: >0.8

	Australia	Belgium (Dutch)	Belgium (French)	China	Denmark	Finland	Germany, Austria and Switzerland	Iceland	India	Latin America	Malaysia
N Managerial	**1,014**	**281**	**290**	**463**	**920**	**1,041**	**460**	**147**	**301**	**331**	**397**
N Non-Managerial	**991**	**1,078**	**829**	**378**	**2,137**	**2,110**	**467**	**273**	**120**	**127**	**610**
Persuasive	0.32	0.47	0.33	0.04	0.44	0.54	0.35	0.47	0.12	0.19	0.12
Controlling	0.70	0.80	0.69	0.46	1.00	1.00	0.86	0.71	0.46	0.26	0.63
Outspoken	0.34	0.22	0.33	0.23	0.21	0.40	0.28	0.20	0.07	-0.06	0.30
Independent Minded	0.10	0.21	0.25	0.07	0.14	0.37	0.01	0.24	-0.04	0.17	0.02
Outgoing	0.05	-0.03	-0.06	-0.17	0.05	0.16	-0.16	0.04	-0.22	-0.10	-0.13
Affiliative	-0.10	-0.30	-0.13	-0.28	-0.11	-0.14	-0.09	-0.19	-0.01	-0.13	0.07
Socially Confident	0.05	0.19	0.10	-0.19	0.17	0.36	0.05	0.04	0.00	-0.03	0.02
Modest	-0.10	-0.10	-0.19	0.06	-0.02	-0.15	-0.11	-0.11	0.01	0.10	-0.12
Democratic	0.17	0.14	0.21	-0.07	0.26	0.02	0.08	0.04	0.07	-0.32	0.15
Caring	-0.02	-0.06	-0.07	0.02	0.04	-0.21	-0.20	-0.22	0.14	-0.24	0.07
Data Rational	-0.06	0.14	0.30	0.08	0.06	0.10	0.13	0.09	0.31	-0.06	0.07
Evaluative	0.21	0.32	0.51	0.23	0.22	0.49	0.15	0.10	0.12	-0.02	0.29
Behavioural	0.19	0.13	0.22	0.03	0.35	0.20	-0.08	0.02	0.03	0.14	0.07
Conventional	-0.30	-0.46	-0.12	-0.06	-0.36	-0.39	-0.16	-0.12	0.03	-0.26	-0.04
Conceptual	0.06	0.14	0.24	0.04	-0.03	0.04	-0.07	-0.11	-0.17	-0.03	0.17
Innovative	0.24	0.54	0.38	0.20	0.23	0.43	0.32	0.18	0.11	0.00	0.35
Variety Seeking	0.26	0.35	0.30	0.20	0.29	0.33	0.23	-0.01	-0.01	0.21	0.08
Adaptable	0.04	0.00	-0.01	-0.21	-0.09	-0.04	-0.12	0.05	-0.11	0.15	-0.10
Forward Thinking	0.24	0.42	0.42	0.16	0.61	0.38	0.28	0.19	0.35	0.00	0.23
Detail Conscious	-0.15	-0.42	-0.21	-0.07	-0.11	-0.21	-0.27	-0.20	-0.04	-0.26	0.10
Conscientious	-0.13	-0.32	-0.34	0.08	-0.05	-0.26	-0.25	-0.22	0.08	-0.15	0.24
Rule Following	-0.32	-0.32	-0.28	0.07	-0.19	-0.47	-0.28	0.01	-0.14	-0.28	-0.06
Relaxed	-0.09	-0.17	-0.04	-0.02	0.01	-0.13	0.12	-0.21	-0.19	-0.22	-0.23
Worrying	-0.18	-0.37	-0.34	-0.10	-0.33	-0.45	-0.35	-0.27	-0.21	0.01	-0.22
Tough Minded	0.05	0.22	0.08	-0.15	0.19	0.15	0.14	-0.06	-0.07	-0.03	0.05
Optimistic	0.17	0.05	0.06	0.02	0.25	0.16	0.22	-0.05	-0.04	-0.13	0.16
Trusting	0.19	0.14	0.28	0.05	0.15	-0.03	0.07	-0.06	0.32	0.06	0.25
Emotionally Controlled	-0.13	0.05	-0.09	-0.05	-0.08	-0.03	0.06	-0.15	-0.22	0.19	-0.23
Vigorous	-0.11	-0.24	-0.23	-0.17	-0.03	-0.18	-0.23	-0.16	-0.11	-0.07	-0.08
Competitive	0.07	0.19	0.14	0.12	0.34	0.46	0.30	0.31	0.29	0.22	0.07
Achieving	0.17	0.32	0.17	0.06	0.52	0.52	0.19	0.20	0.15	0.02	0.09
Decisive	0.46	0.44	0.48	0.43	0.47	0.56	0.44	0.56	0.36	0.44	0.21

	Middle East	Netherlands	New Zealand	Norway	Poland	Portugal	Singapore	South Africa	Sweden	UK	US
N Managerial	151	962	149	1,144	231	106	411	824	2,143	4,618	749
N Non-Managerial	118	2,010	167	1,404	193	263	509	563	5,100	4,912	1,561
Persuasive	-0.06	0.41	0.29	0.34	0.28	0.63	-0.08	0.06	0.39	0.08	0.03
Controlling	0.29	0.84	0.58	0.81	0.61	0.78	0.52	0.58	0.71	0.52	0.38
Outspoken	0.22	0.25	0.29	0.28	0.05	0.34	0.24	0.26	0.21	0.28	0.18
Independent Minded	0.10	0.20	0.09	0.00	0.19	0.34	0.19	0.12	0.08	0.31	0.04
Outgoing	-0.15	0.12	-0.10	-0.02	-0.15	0.08	-0.41	-0.09	0.02	-0.04	-0.06
Affiliative	-0.12	-0.16	-0.27	-0.11	-0.44	-0.15	-0.17	-0.17	-0.20	-0.17	-0.11
Socially Confident	0.13	0.27	-0.01	0.18	-0.08	0.22	-0.17	-0.03	0.01	-0.16	-0.05
Modest	-0.14	-0.16	0.09	0.02	0.10	0.02	0.02	0.01	-0.04	0.06	0.16
Democratic	-0.15	0.15	-0.01	0.35	-0.04	0.03	0.04	0.18	0.20	0.21	0.18
Caring	0.16	0.02	-0.01	-0.01	0.05	-0.08	-0.04	-0.17	0.05	-0.19	0.03
Data Rational	0.05	0.02	-0.03	0.24	0.09	0.14	0.06	0.05	0.00	-0.07	-0.17
Evaluative	0.10	0.21	-0.01	0.30	0.17	0.19	0.38	0.25	0.15	0.13	0.02
Behavioural	0.03	0.14	0.15	0.14	-0.08	0.28	-0.01	0.01	0.28	0.12	0.12
Conventional	-0.02	-0.30	-0.32	-0.18	-0.04	-0.28	-0.15	-0.26	-0.18	-0.38	-0.14
Conceptual	-0.06	0.02	-0.01	-0.03	-0.11	-0.30	-0.06	0.05	-0.07	-0.04	-0.03
Innovative	0.12	0.40	-0.03	0.34	0.33	-0.02	0.22	0.20	0.23	0.16	0.17
Variety Seeking	-0.22	0.24	0.29	0.21	-0.04	0.33	0.32	0.26	0.11	0.43	0.17
Adaptable	-0.18	-0.03	0.09	-0.23	0.03	0.15	-0.08	-0.04	-0.17	0.06	-0.03
Forward Thinking	0.06	0.34	0.21	0.48	0.27	0.15	0.22	0.23	0.40	0.19	0.04
Detail Conscious	-0.07	-0.28	-0.21	0.00	-0.03	-0.18	-0.08	-0.30	-0.20	-0.39	-0.25
Conscientious	0.11	-0.08	-0.33	-0.01	0.00	-0.05	0.08	-0.21	-0.09	-0.35	-0.22
Rule Following	-0.11	-0.29	-0.30	-0.01	0.00	-0.15	-0.03	-0.28	-0.14	-0.48	-0.17
Relaxed	0.01	0.07	-0.23	-0.01	-0.11	-0.21	-0.11	-0.07	0.00	-0.15	-0.09
Worrying	-0.04	-0.41	-0.17	-0.35	-0.30	-0.29	-0.07	-0.23	-0.26	-0.05	-0.13
Tough Minded	-0.05	0.22	0.06	0.21	0.05	0.02	-0.14	0.12	0.13	-0.09	-0.10
Optimistic	-0.06	0.25	0.13	0.13	0.07	0.01	0.07	-0.01	0.21	0.09	0.14
Trusting	0.47	0.15	-0.06	0.17	0.09	0.03	0.35	0.23	0.21	0.14	0.26
Emotionally Controlled	-0.05	-0.19	0.03	0.01	0.29	-0.14	-0.09	-0.01	-0.08	0.03	-0.03
Vigorous	-0.11	-0.18	-0.20	0.00	-0.24	0.11	-0.12	-0.27	-0.26	-0.35	-0.24
Competitive	0.08	0.25	0.41	0.22	0.37	0.58	-0.09	0.08	0.27	0.13	-0.03
Achieving	0.17	0.36	0.10	0.37	0.16	0.36	0.03	0.06	0.24	-0.03	-0.06
Decisive	0.25	0.49	0.35	0.46	0.49	0.51	0.38	0.37	0.51	0.36	0.40

Figure 12: Differences on Controlling comparing managers and non-managers (positive effect sizes indicate managers score higher than non-managers)
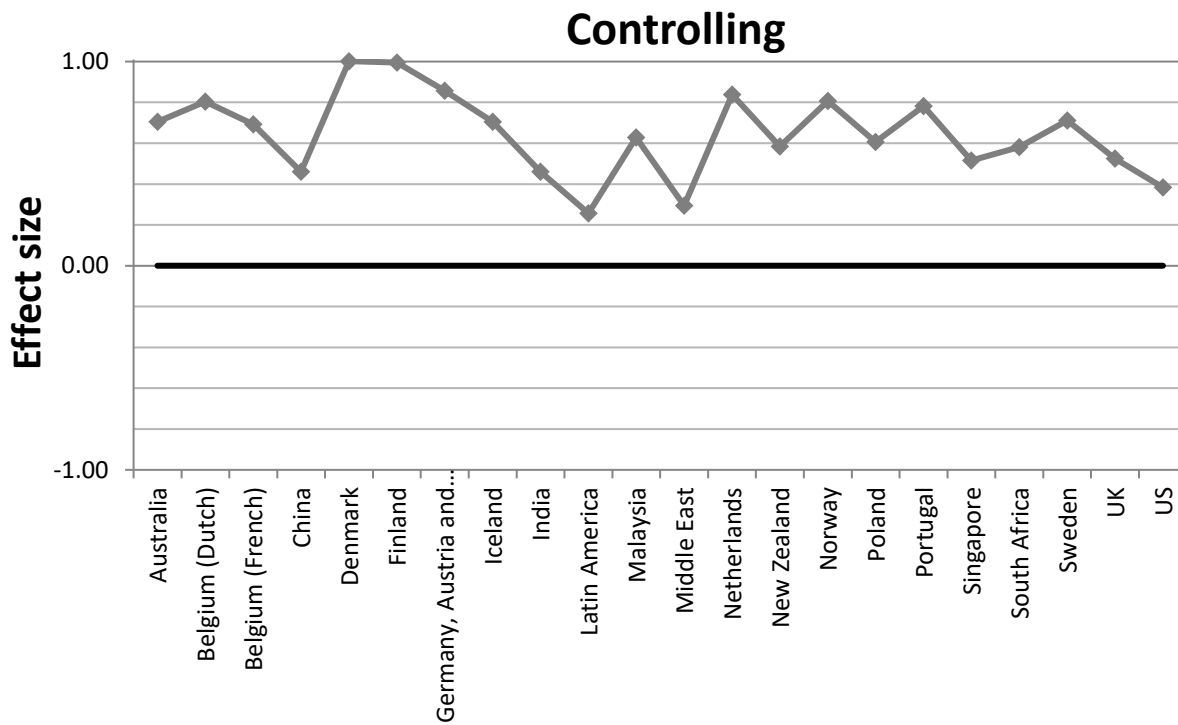
Figure 13: Differences on Persuasive comparing managers and non-managers (positive effect sizes indicate managers score higher than non-managers)
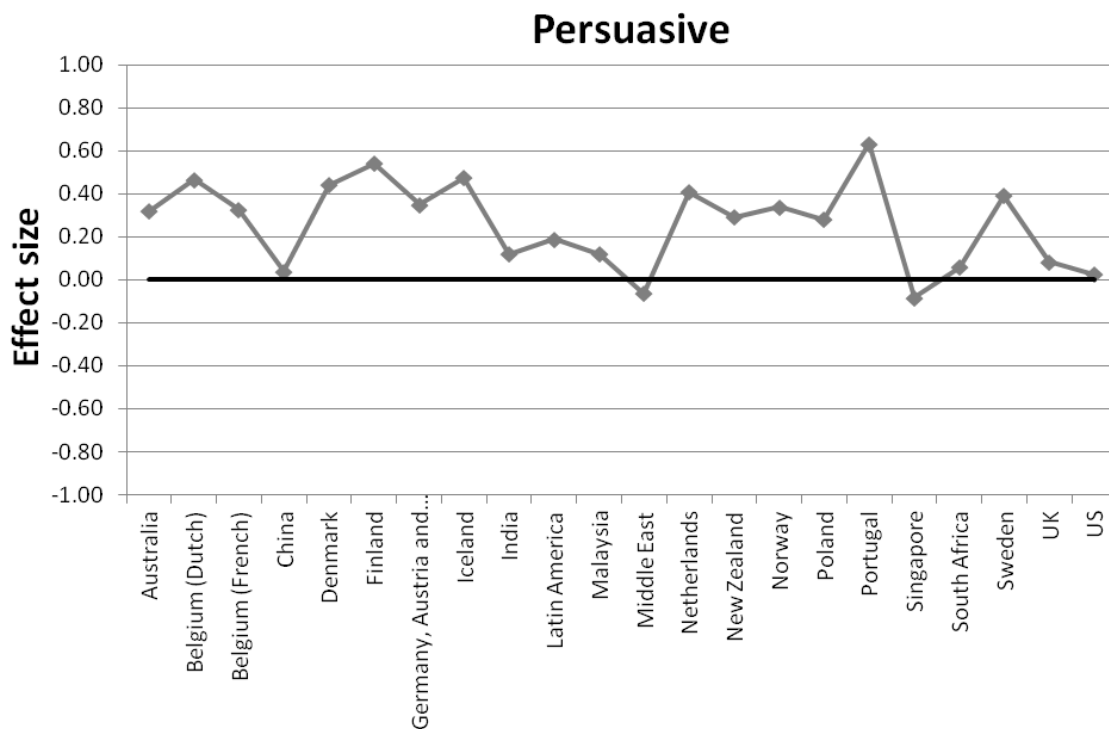
Figure 14: Differences on Conscientious comparing managers and non-managers (positive effect sizes indicate managers score higher than non-managers)
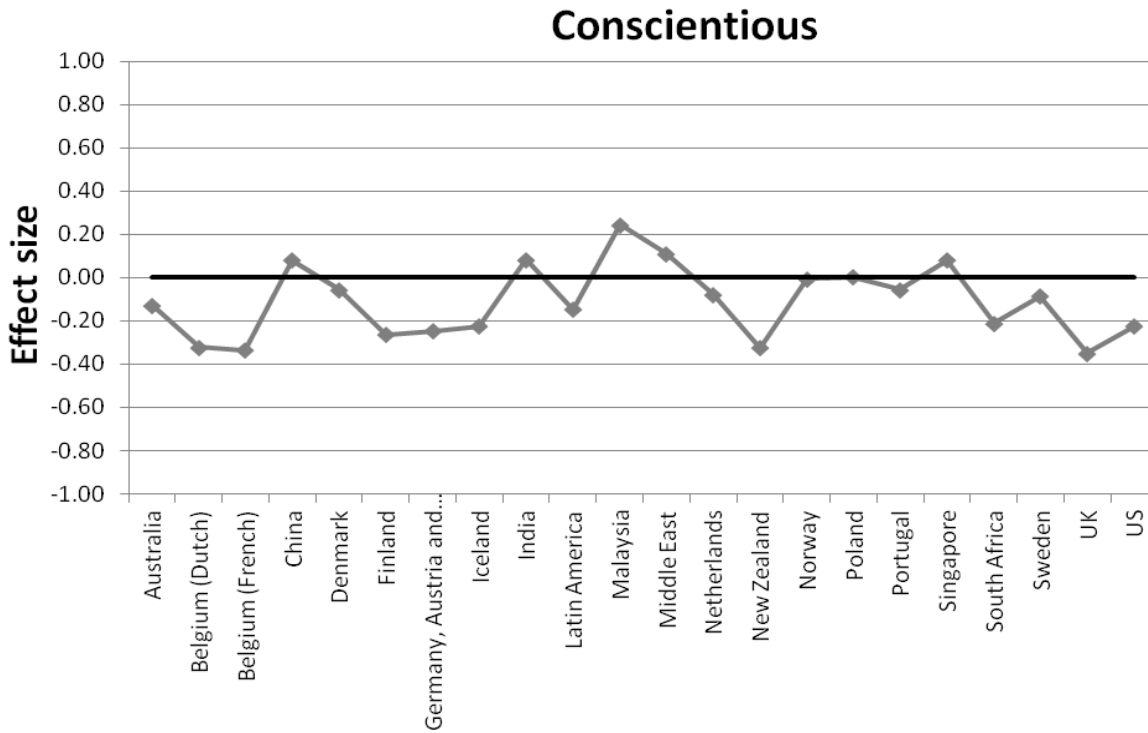
Figure 15: Differences on Competitive comparing managers and non-managers (positive effect sizes indicate managers score higher than non-managers)
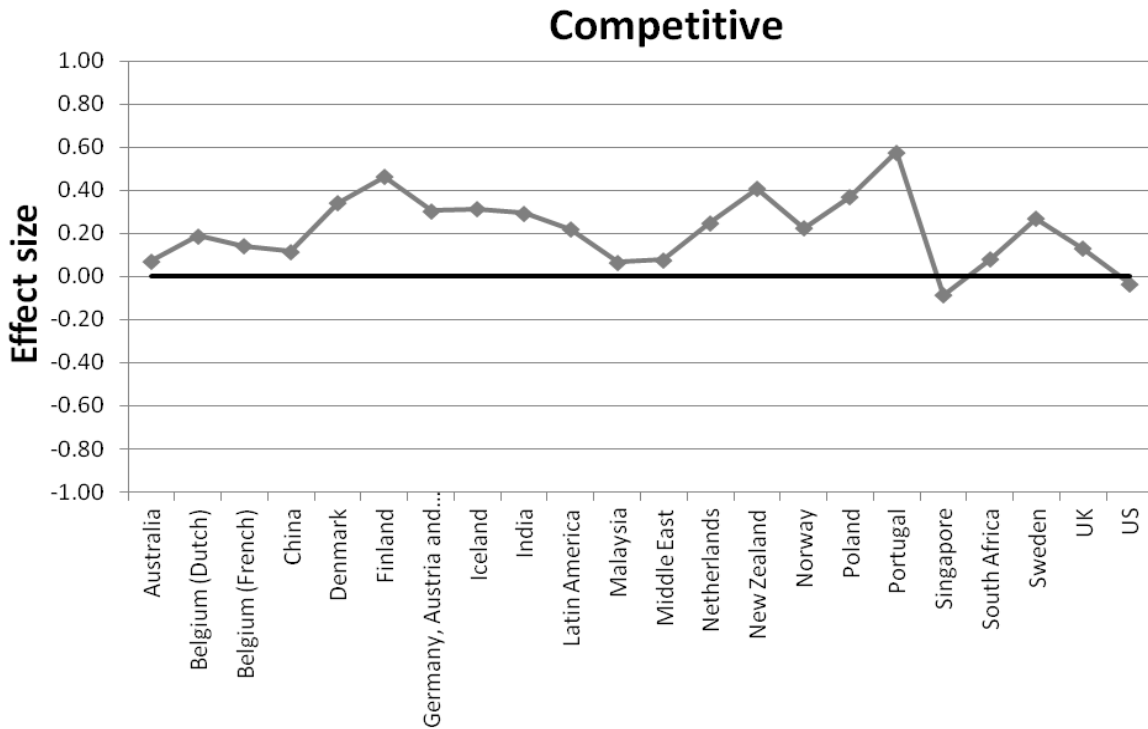
6.6 Additional Group Comparisons

Due to local demands in the US, additional adverse impact analysis was completed in 2014. The group differences of Ethnicity, Age and Gender were examined.

This sample included 3195 US candidates who completed the instrument between January 2011 and April 2014. As reporting demographic information is optional, only those who reported at least some demographic data were included in this analysis. Sample sizes used to calculate subgroup difference effect sizes are provided in Table 41. Effect sizes for subgroup comparisons were not reported when the subgroup sample size (N) was less than 200 because samples smaller than this are more susceptible to sampling error and typically lack the statistical power necessary to detect differences at critical d thresholds (Cohen, 1988). Therefore, results for the Hispanic or Latino group are not reported in this addendum. Statistics for other racial/ethnic groups where data were not available will be updated upon collection of sufficient data.

Table 41: US 2014 Adverse Impact Analysis Sample Demographics

Group	Gender		Age		Racial/ Ethnic Group			
Subgroup	Male	Female	<40 years	≥40 years	White	Asian	Black/African American	Hispanic or Latino
N	1,751	689	995	1,462	1,248	1,278	531	138

Subgroup difference effect sizes were calculated on mean OPQ32r theta scores using the d statistic (see Table 42); the standardized mean score difference between groups for each of the scales of the OPQ32r. The effect size statistic (d) is simply the average score difference between groups, in standard deviation units. Across contexts, a d value of|0.2 is considered small, 0.5 is considered medium, and 0.8 is considered large (Cohen, 1988).

The subgroup difference analysis presented here indicated minimal differences between groups. Effects indicating a difference in scores between groups, when found, were generally small according to well-established professional guidelines for interpreting effect sizes. The sole exception for race/ethnicity was the difference in scores among the White and Black samples, where a moderate effect was identified such that Blacks scored higher than Whites on the scale 'Persuasive'. The minor exceptions for gender were moderate effects where females scored higher than males on the scale 'Detail Conscious' and males scored higher than females on the 'Competitive' scale. These results are consistent with established research indicating minimal differences amongst race/ethnic, age and gender groups using personality instruments Hough, 1998; Hough et al., 2001; Ones & Anderson, 2002; Schmitt, Clause, & Pulakos, 1996).

Table 42: US 2014 Adverse Impact Analysis: Effect Size

Effect sizes are highlighted as following:

Effect Sizes		
Small: >|0.2| to ≤|0.5|		
Medium: >| 0.5| to ≤ |0.8|		

OPQ32r Scale	Female*	≥40**	Asian***	Black/ African American***
Persuasive	-0.24	0	0.11	0.52
Controlling	-0.25	0.16	0.04	0.27
Outspoken	-0.14	-0.05	0	-0.14
Independent Minded	-0.05	0.01	0.06	0.01
Outgoing	0.14	-0.23	-0.08	0.18
Affiliative	0.23	-0.17	-0.08	-0.07
Socially Confident	0.02	-0.15	-0.02	0.25
Modest	-0.04	0.28	0.09	-0.1
Democratic	-0.07	0.1	0.02	-0.17
Caring	0.36	-0.14	-0.11	-0.13
Data Rational	-0.15	-0.18	-0.02	-0.13
Evaluative	-0.09	-0.19	0.01	-0.09
Behavioral	0.22	-0.14	-0.01	0.1
Conventional	0.08	0.02	-0.06	0.01
Conceptual	-0.06	-0.19	0.07	-0.04
Innovative	-0.18	-0.02	0.06	0.06
Variety Seeking	-0.1	0.12	-0.02	-0.11
Adaptable	0.08	-0.12	-0.05	0.03
Forward Thinking	0	-0.04	0	0.1
Detail Conscious	0.51	-0.27	-0.06	0.06
Conscientious	0.27	-0.13	0.01	0.09
Rule Following	0.11	-0.05	-0.08	0.15
Relaxed	-0.18	-0.17	0	0.04
Worrying	0.38	0.01	-0.06	-0.27
Tough Minded	-0.33	-0.1	0.1	0.11
Optimistic	0.18	0.07	-0.07	0.08
Trusting	-0.04	0.38	0.06	-0.23
Emotionally Controlled	-0.08	0.09	0.05	0.01
Vigorous	0.5	-0.16	-0.05	-0.04
Competitive	-0.62	0	-0.01	0.48
Achieving	0.03	-0.31	-0.05	0.25
Decisive	-0.14	0.24	0.06	-0.09

Chapter 7: Comparisons by country or region

Comparisons by country or regions are only meaningful once both construct and scalar equivalence have been established. Scalar equivalence represents the highest level of equivalence where raw scores can be directly compared across groups (Van de Vijver & Poortinga, 2005). When there is scalar equivalence across two groups, it means that a given raw score represents the same amount of the measured trait regardless of which group the person comes from. In such a case, when norming raw scores it is more appropriate to use an aggregate norm across the two groups than to use individual group norms. The latter would effectively remove the raw scores difference between the groups, removing a real effect and introducing a systematic bias.

The research findings obtained in the last few years (from around 2007 onwards) support the view that differences between countries or regions based on OPQ32 average raw scale scores represent real differences in people's preferred ways of behaving, thinking and feeling rather than bias in measurement (e.g. introduced through translation) or sampling bias.

Bartram (2013) demonstrated the scalar equivalence of the OPQ32 constructs through Big Five scale scores, derived from OPQ32i data, for a sample of over one million people in terms of differences between 31 countries involving over 20 different languages. Strong relationships were found between country average scale scores and country standard deviations (SDs), and several independent global measures such as Hofstede's cultural dimensions and country-level performance indicators (the World Economic Forum Global Competitiveness Index and UN Human Development indices). Country SDs were observed to vary with cultural "tightness" ratings. Strong correlations were found between these indicators and both country-level mean personality scores and SDs of personality scores, supporting the argument that between-country differences represent true score variance rather than systematic instrument-related biases. Issues to consider when testing candidates across countries are discussed in further detail in Bartram (2012). Practical guidance on when and how to apply international norms when testing candidates across different languages or language versions of the OPQ32r are provided in the documentation for the OPQ32r international norm (SHL, 2015b) which can be obtained free of charge by contacting an account manager.

Using Sample 6 described in section 2.6, the following section describes scale score differences across countries or regions. Aggregate (mean) personality scores are given for each of the regions. Means and standard deviations for each of the OPQ scales by country or region can be found in Table 43. Mean scores of individual countries or regions were compared against the overall sample mean based on sten scores (mean=5.5, SD=2) and converted into effect sizes, which can be found in Table 44.

Effect size conventions established by Cohen (1988) suggest that effects of 0.2 (0.4 sten) are small, 0.5 (1 sten) are medium and 0.8 (1.6 sten) are large. For practical applications a medium effect is defined here as equal to or larger than 0.5 sten (0.25 d score) and a large effect as equal to or exceeding 1.5 sten (0.75 d score). Rounding these stens would result in 1 sten for a medium effect and 2 stens for a large effect. Effects smaller than 0.5 sten are considered as having little practical impact.

Results show that mean scale score deviations from the overall average were small, typically falling within 0.2 of a standard deviation. Most differences were as expected; countries that are culturally similar yielded more similar OPQ32 profiles. Aggregate personality scores across countries and effect sizes are displayed graphically in Figures 16 to 47. Mean personality scores are provided on the left and effect sizes on the right.

From these results we can see that on the Vigorous scale Greater China has the lowest aggregate (mean) score (effect size -0.71), whereas Iceland has the highest (0.67). Both the UK (0.43) and US (0.28) score higher on the Vigorous scale compared to the overall sample (small effect size). With regards to the OPQ scale Adaptable, Turkey has the lowest aggregate score (-0.63) and Greater China has the highest (0.77), while the US (-0.32) shows a small difference and UK (0.07) hardly any compared to the overall sample mean. For Trusting, Turkey shows the lowest aggregate score (-0.67) and Denmark the highest (0.76) whereas the UK (-0.10) and UK (-0.08) show very little difference from the overall mean score.

Table 43: Mean score and standard deviation per country or region.

	Australia		Belgium (Dutch)		Belgium (French)		Brazil		Canada	
N	9,120		2,385		2,529		1,006		703	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Persuasive	5.47	2.07	5.18	2.08	4.96	1.92	5.84	1.84	5.51	1.96
Controlling	5.36	1.92	4.46	2.17	4.73	2.14	6.20	1.42	5.93	1.84
Outspoken	4.90	1.97	5.66	2.12	5.39	1.95	5.49	1.89	4.96	1.96
Independent Minded	5.01	1.97	5.33	2.14	5.17	1.98	5.09	1.83	5.11	1.99
Outgoing	5.37	2.11	5.53	2.10	5.13	1.75	5.33	1.91	5.57	2.02
Affiliative	5.54	1.97	6.15	2.12	5.75	1.92	5.15	1.69	5.55	1.83
Socially Confident	5.79	2.11	5.20	2.18	5.43	1.87	4.86	1.61	5.98	2.09
Modest	5.81	2.06	5.25	1.99	6.13	1.90	5.16	1.91	5.78	2.06
Democratic	5.71	2.08	6.38	1.91	5.86	1.90	5.38	1.74	5.68	2.06
Caring	5.81	2.09	5.73	2.14	4.91	1.99	5.42	1.77	6.13	2.10
Data Rational	5.42	2.05	5.29	2.11	5.57	1.69	5.87	1.73	5.86	2.22
Evaluative	5.50	2.01	6.02	1.93	5.87	2.11	5.75	1.71	5.80	1.92
Behavioural	5.71	2.09	5.58	2.22	5.19	2.15	5.88	1.75	6.10	2.07
Conventional	5.93	2.06	5.82	2.14	5.42	1.93	5.39	1.75	5.55	2.02
Conceptual	5.92	2.07	5.68	2.10	5.78	1.94	5.36	1.90	5.99	2.00
Innovative	5.43	1.95	5.18	2.10	4.86	1.96	5.18	1.76	5.86	1.97
Variety Seeking	5.05	2.03	5.85	2.31	5.63	2.03	5.50	1.85	5.43	2.13
Adaptable	5.37	2.02	4.97	2.05	6.25	1.66	5.26	1.76	5.22	1.86
Forward Thinking	5.53	1.95	4.81	2.14	5.19	1.89	5.38	1.78	5.89	2.09
Detail Conscious	5.76	2.06	5.06	2.18	5.89	1.88	4.89	2.02	5.69	2.05
Conscientious	5.66	2.00	5.67	2.02	5.78	1.92	4.66	1.77	5.63	1.95
Rule Following	5.48	2.06	5.07	1.95	5.71	2.00	5.02	1.78	5.38	2.06
Relaxed	5.52	2.10	6.12	2.20	5.16	2.02	5.07	1.61	5.16	2.10
Worrying	5.36	2.02	5.45	2.03	5.77	1.99	6.75	1.55	5.29	1.96
Tough Minded	5.70	2.14	5.45	2.09	4.83	1.87	5.13	1.91	5.84	2.10
Optimistic	5.51	1.94	5.15	2.02	4.44	2.01	5.55	1.71	5.51	2.01
Trusting	5.46	1.83	5.29	1.98	5.12	1.91	5.53	1.82	5.70	1.84
Emotionally Controlled	5.50	1.95	5.15	2.25	6.13	1.92	4.92	1.86	5.39	2.02
Vigorous	6.14	1.84	6.06	1.74	5.31	1.81	4.71	1.76	6.03	1.89
Competitive	5.09	2.09	4.49	2.04	4.83	1.77	5.86	1.67	5.42	2.07
Achieving	5.54	1.95	5.11	1.99	4.69	1.87	5.38	1.81	5.95	2.06
Decisive	4.95	1.94	5.37	2.01	5.29	1.94	6.11	1.91	4.97	1.91

	China		Denmark		Finland		France		Germany, Austria and Switzerland	
N	3,322		6,809		5,381		4,225		2,336	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Persuasive	5.05	1.72	5.91	2.14	5.87	2.13	5.60	1.96	5.82	1.88
Controlling	4.99	1.89	5.93	2.17	5.04	2.28	5.45	2.06	5.85	1.63
Outspoken	4.40	1.74	6.09	1.81	5.13	2.03	5.40	1.83	5.83	1.88
Independent Minded	6.44	1.62	5.81	1.89	5.55	2.09	5.16	1.91	6.18	1.87
Outgoing	5.33	2.19	6.54	1.96	6.23	2.21	5.12	1.67	5.35	1.83
Affiliative	4.75	1.88	6.34	1.69	5.46	1.99	5.79	1.71	5.56	1.94
Socially Confident	5.20	2.09	6.32	1.90	5.89	2.20	5.50	1.74	6.02	1.83
Modest	5.79	1.92	4.84	1.85	5.11	2.00	6.07	1.77	5.13	1.91
Democratic	5.95	1.73	5.83	2.05	5.75	1.96	5.73	1.82	5.78	1.88
Caring	6.13	1.63	6.04	1.92	5.42	1.93	5.10	1.88	5.96	1.85
Data Rational	5.20	1.75	5.26	1.90	4.69	1.91	5.73	1.72	4.76	1.63
Evaluative	4.78	1.82	5.99	1.93	5.20	2.05	5.96	1.98	5.48	1.69
Behavioural	5.18	1.76	6.05	1.92	4.83	2.06	5.79	1.96	5.30	1.90
Conventional	5.78	1.85	5.17	2.00	5.76	2.02	5.00	1.83	4.31	1.81
Conceptual	5.44	1.64	5.51	2.09	4.73	2.12	5.49	1.98	5.63	1.91
Innovative	5.26	2.00	5.55	2.06	4.92	2.08	4.98	1.95	5.94	1.79
Variety Seeking	4.72	1.77	5.52	1.95	5.06	1.89	6.11	1.92	6.13	1.76
Adaptable	6.68	1.79	5.40	1.88	5.33	2.03	6.56	1.60	5.75	1.70
Forward Thinking	6.40	1.82	5.35	1.98	5.02	2.11	5.29	1.86	5.10	1.87
Detail Conscious	5.48	1.75	5.81	1.94	5.10	1.81	5.59	1.93	5.34	1.59
Conscientious	5.31	1.60	5.96	2.01	5.47	1.94	5.57	1.86	5.64	1.89
Rule Following	3.53	1.32	4.96	1.86	5.95	1.93	5.47	1.84	5.42	1.77
Relaxed	5.25	1.84	6.36	1.69	6.20	1.95	5.10	1.85	5.81	1.85
Worrying	5.59	1.88	4.82	1.90	5.29	2.10	5.59	1.92	5.12	2.02
Tough Minded	5.86	1.81	6.57	1.85	5.55	1.94	4.70	1.89	5.74	1.91
Optimistic	5.59	1.97	6.41	1.78	6.14	2.08	4.84	1.97	5.98	2.01
Trusting	6.22	1.85	7.02	1.62	6.27	1.83	5.30	1.76	5.20	1.79
Emotionally Controlled	5.54	2.03	5.32	1.81	5.14	2.09	6.53	1.77	5.51	1.97
Vigorous	4.47	1.94	6.49	1.75	5.59	1.93	5.15	1.69	5.28	1.74
Competitive	6.15	1.45	5.69	2.17	4.77	2.17	5.22	2.01	4.84	1.73
Achieving	4.56	1.77	5.87	1.97	5.15	2.17	5.29	1.85	5.96	1.74
Decisive	4.62	2.05	5.82	1.85	6.28	1.95	5.66	1.99	5.55	1.59

	Greater China		Iceland		India		Italy		Korea	
N	**564**		**729**		**1,098**		**4,550**		**846**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Persuasive	4.97	1.72	5.42	1.95	5.37	1.96	4.89	1.86	5.28	1.89
Controlling	4.43	1.94	5.37	1.90	5.80	1.83	5.57	1.96	4.25	1.51
Outspoken	4.27	1.92	5.54	1.75	5.36	1.87	6.01	2.02	5.41	1.88
Independent Minded	5.92	1.74	6.08	1.74	6.01	2.03	5.90	1.84	5.04	2.19
Outgoing	5.75	2.03	5.02	1.79	4.91	1.80	5.81	1.91	5.84	1.98
Affiliative	4.56	1.98	5.75	1.94	4.80	1.98	5.22	1.96	6.43	2.21
Socially Confident	4.90	1.97	5.41	1.68	5.29	1.94	4.94	1.95	5.45	2.00
Modest	5.91	2.01	5.95	1.84	5.85	2.05	5.82	2.04	4.79	1.97
Democratic	5.87	1.75	5.79	1.94	4.87	1.97	4.81	1.85	6.04	1.83
Caring	5.72	1.83	5.68	2.02	5.25	1.92	4.91	1.93	5.11	1.88
Data Rational	5.20	1.92	5.50	1.98	6.15	2.00	5.15	1.81	5.03	2.06
Evaluative	4.37	2.04	5.83	1.87	5.08	1.80	5.81	1.78	3.37	1.79
Behavioural	5.53	1.84	4.92	2.15	5.00	1.86	5.54	1.94	5.04	1.62
Conventional	5.55	1.92	5.30	1.90	5.79	2.12	5.18	2.17	5.47	1.88
Conceptual	6.10	1.80	5.14	1.99	5.25	1.64	4.65	1.84	4.93	1.99
Innovative	4.72	2.04	5.66	1.97	5.75	1.92	5.26	1.91	4.58	2.00
Variety Seeking	4.99	2.07	6.44	1.57	5.12	2.04	5.92	2.02	4.95	1.83
Adaptable	7.05	1.75	4.57	1.75	5.68	1.97	5.72	1.98	6.72	1.44
Forward Thinking	5.80	1.91	6.03	1.85	6.05	2.07	4.74	1.83	4.92	1.90
Detail Conscious	4.96	1.86	5.69	2.11	5.48	1.98	4.62	1.82	4.85	1.80
Conscientious	4.85	1.71	5.53	2.09	5.41	1.90	5.15	1.80	4.09	1.97
Rule Following	4.10	1.50	5.97	1.99	6.09	2.01	5.26	1.91	5.96	1.76
Relaxed	4.82	1.76	6.11	1.93	5.35	2.01	5.06	1.94	4.93	1.65
Worrying	5.90	1.91	4.90	1.71	5.16	1.96	5.98	1.80	6.06	1.93
Tough Minded	5.91	1.82	6.17	1.84	4.97	1.78	5.24	2.06	5.04	1.84
Optimistic	4.89	2.02	6.12	1.87	6.00	1.90	4.58	1.96	5.09	2.01
Trusting	5.01	1.91	6.57	1.66	5.39	2.15	4.97	2.08	6.19	1.94
Emotionally Controlled	5.85	2.10	5.36	1.86	5.92	1.78	5.72	1.98	5.99	2.10
Vigorous	4.08	1.81	6.84	1.87	5.40	1.97	5.09	1.89	4.33	1.92
Competitive	5.58	1.78	4.74	2.08	6.09	1.62	5.71	1.82	5.05	1.58
Achieving	4.35	1.91	5.81	1.82	5.90	1.91	5.09	2.00	4.15	1.85
Decisive	4.77	2.10	6.29	1.95	5.31	1.83	6.19	1.95	4.63	1.97

SHL.

	Latin America		Malaysia		Middle East		Netherlands		New Zealand	
N	**971**		**2,672**		**1,494**		**6,227**		**2,818**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Persuasive	6.41	1.90	5.08	1.77	5.05	1.47	5.09	2.05	5.16	1.99
Controlling	6.62	1.70	4.87	1.82	5.33	1.76	5.06	2.14	5.29	1.92
Outspoken	6.74	1.72	5.01	1.80	5.70	1.54	5.99	2.12	5.12	2.01
Independent Minded	5.30	1.75	5.47	1.93	6.06	1.83	6.42	1.96	4.84	1.97
Outgoing	5.35	1.74	5.47	1.88	4.91	1.44	5.78	2.01	5.27	2.09
Affiliative	4.86	1.82	4.74	1.76	4.64	1.89	6.61	1.90	5.45	1.97
Socially Confident	5.96	1.39	5.15	1.83	4.26	1.26	5.64	2.05	5.66	2.07
Modest	5.34	1.92	5.83	1.91	5.88	1.67	5.25	1.99	5.78	2.01
Democratic	4.27	1.86	5.43	1.78	5.09	1.79	5.80	1.86	5.73	2.10
Caring	5.27	1.90	4.92	1.82	4.97	1.70	6.18	2.03	5.55	2.05
Data Rational	6.52	1.73	5.37	1.87	5.55	1.75	5.12	2.07	5.49	2.10
Evaluative	5.72	1.61	4.39	1.86	4.47	1.68	5.89	1.93	5.58	2.03
Behavioural	5.30	1.69	5.38	1.78	4.47	1.64	6.20	2.05	5.38	2.11
Conventional	5.04	1.74	6.20	1.77	6.18	1.76	4.93	1.99	6.11	2.13
Conceptual	5.42	1.74	5.63	1.73	5.36	1.68	5.88	2.04	5.89	2.11
Innovative	6.28	1.93	4.75	1.77	5.54	1.53	5.43	2.04	5.47	1.96
Variety Seeking	5.65	1.92	4.86	1.94	5.09	1.67	6.30	2.06	4.92	2.02
Adaptable	4.27	1.88	6.23	1.87	5.97	1.56	5.28	1.85	5.20	2.01
Forward Thinking	5.89	1.76	5.82	1.85	5.67	1.79	4.83	2.03	5.45	1.98
Detail Conscious	5.61	1.87	5.35	1.93	5.87	1.69	4.94	2.03	5.89	2.08
Conscientious	5.89	1.78	4.74	1.85	5.59	1.85	5.02	1.93	5.77	2.01
Rule Following	5.79	1.81	5.51	1.92	6.93	1.73	4.62	1.95	5.44	2.10
Relaxed	4.87	1.76	5.19	2.02	5.00	1.58	6.60	2.11	5.83	2.08
Worrying	5.55	1.71	5.73	1.80	5.27	1.69	5.09	1.95	5.36	2.01
Tough Minded	5.57	1.77	5.02	1.84	5.54	1.70	6.02	2.01	5.67	2.10
Optimistic	5.43	1.63	5.29	1.89	5.32	1.67	5.67	2.01	5.52	1.93
Trusting	5.26	1.89	4.65	1.84	4.71	1.68	5.91	1.82	5.57	1.83
Emotionally Controlled	5.03	1.79	5.90	1.89	6.33	1.62	5.16	2.21	5.49	1.97
Vigorous	5.70	1.81	4.31	1.88	4.56	1.81	6.02	1.68	6.32	1.80
Competitive	6.09	1.71	5.62	1.65	6.46	1.43	4.92	2.13	5.10	2.11
Achieving	6.22	1.74	5.08	1.83	5.60	1.68	5.16	2.14	5.46	1.92
Decisive	6.46	1.77	4.90	1.73	4.77	1.72	5.99	1.93	5.15	2.00

SHL.

	Norway		Poland		Portugal		Singapore		South Africa	
N	**4,659**		**927**		**2,248**		**4,053**		**4,880**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Persuasive	5.70	2.01	5.43	1.87	5.06	2.00	5.08	2.10	5.39	1.86
Controlling	6.03	2.00	5.45	1.91	5.39	1.99	5.19	2.08	5.53	1.81
Outspoken	6.10	1.89	6.00	1.98	5.58	1.93	4.97	1.96	5.47	1.98
Independent Minded	4.68	2.01	6.23	1.91	5.05	1.80	5.06	1.96	5.67	1.88
Outgoing	6.06	2.02	4.90	1.81	4.99	1.90	5.30	2.07	5.14	1.90
Affiliative	6.35	1.73	4.81	1.93	5.39	1.80	5.10	1.96	4.48	1.82
Socially Confident	6.16	1.85	4.90	2.08	5.10	1.88	5.43	2.16	5.20	1.78
Modest	5.18	1.78	5.57	1.93	5.39	1.90	5.91	2.11	5.63	1.88
Democratic	6.36	2.07	4.72	1.87	5.85	1.82	5.76	1.94	5.02	1.90
Caring	5.78	1.96	4.57	2.00	5.57	1.81	5.46	2.02	5.12	1.94
Data Rational	5.42	1.77	4.99	1.97	5.72	2.00	5.61	2.05	5.75	1.98
Evaluative	5.70	2.00	5.74	1.72	6.09	1.84	5.27	2.07	5.29	1.87
Behavioural	5.80	1.82	5.26	2.02	5.72	1.97	5.91	1.94	4.81	1.86
Conventional	5.43	1.91	5.52	1.89	5.40	1.69	5.98	2.06	6.04	1.86
Conceptual	4.98	2.08	5.09	2.00	5.99	1.94	5.76	1.94	5.67	1.75
Innovative	5.64	2.01	5.98	1.90	5.24	1.87	4.95	2.10	5.51	1.74
Variety Seeking	5.84	1.82	5.94	1.85	5.67	1.95	5.23	2.13	5.22	1.86
Adaptable	5.47	1.98	5.32	1.81	5.25	2.06	6.19	2.03	4.77	1.78
Forward Thinking	5.86	2.11	5.12	2.03	5.21	1.86	5.89	2.12	5.66	1.78
Detail Conscious	5.66	1.92	5.56	1.89	5.91	2.09	5.62	2.11	5.86	1.91
Conscientious	6.05	2.04	4.82	1.96	5.62	1.93	5.52	1.97	5.67	1.89
Rule Following	6.09	1.87	5.46	1.92	5.74	1.81	5.15	1.96	5.92	1.90
Relaxed	6.31	1.72	4.83	1.88	5.18	1.93	5.19	1.97	5.38	1.89
Worrying	4.84	1.92	5.92	1.81	6.73	1.75	5.49	2.01	5.54	1.76
Tough Minded	5.44	1.80	5.65	2.12	5.04	1.91	5.53	1.99	5.22	1.97
Optimistic	6.29	1.86	4.97	1.97	4.80	1.98	5.17	2.10	5.52	1.77
Trusting	6.58	1.92	4.79	1.90	5.18	1.93	5.02	1.98	4.51	1.80
Emotionally Controlled	5.15	1.88	5.54	2.02	5.11	1.89	5.69	2.05	5.76	1.76
Vigorous	5.81	1.65	5.57	2.09	5.38	1.82	4.74	2.01	5.57	1.79
Competitive	5.77	2.10	5.37	1.94	5.53	1.77	5.25	1.98	5.65	1.74
Achieving	6.00	1.97	5.28	1.83	5.75	1.86	5.02	2.18	5.83	1.76
Decisive	6.31	1.77	5.93	2.03	5.36	1.95	4.72	1.99	5.26	1.82

	Spain		Sweden		Turkey		UK		US	
N	**695**		**13,244**		**1,107**		**22,612**		**4,114**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Persuasive	6.05	1.99	5.77	1.95	5.72	1.94	5.66	1.95	6.23	1.99
Controlling	5.23	1.72	5.85	1.93	6.23	2.05	5.62	1.96	6.09	1.90
Outspoken	5.11	1.76	5.20	1.80	7.00	1.94	5.21	1.95	4.74	1.89
Independent Minded	5.35	1.87	5.16	1.90	5.82	1.85	5.09	1.98	5.23	1.95
Outgoing	5.42	1.98	6.16	1.88	6.10	1.95	5.45	2.02	5.84	1.95
Affiliative	5.80	1.85	6.57	1.69	6.02	1.93	5.33	1.93	5.51	1.89
Socially Confident	4.53	1.75	6.45	1.87	5.29	1.76	5.98	2.08	6.38	1.99
Modest	5.20	1.85	5.41	1.89	4.65	2.10	5.69	2.04	5.57	1.99
Democratic	4.71	1.87	5.75	2.00	5.25	1.77	5.84	2.05	4.95	2.03
Caring	5.56	1.78	6.38	1.87	4.92	1.89	5.80	2.09	5.67	2.00
Data Rational	5.48	1.87	4.99	1.81	6.82	1.84	5.55	2.12	5.59	2.08
Evaluative	5.80	1.84	5.60	1.95	6.73	1.73	5.91	1.98	5.38	1.92
Behavioural	5.41	1.85	6.06	2.00	6.03	1.81	5.75	2.06	5.69	2.01
Conventional	5.48	1.81	5.51	1.93	4.61	1.83	5.54	2.06	5.78	1.96
Conceptual	5.29	1.97	5.36	2.00	6.14	1.89	5.84	2.06	5.82	2.07
Innovative	5.55	1.81	6.00	1.92	6.73	1.96	5.71	1.97	5.87	1.86
Variety Seeking	5.59	1.87	5.58	1.82	5.95	1.94	5.29	2.03	5.15	1.94
Adaptable	5.46	1.94	6.07	1.93	4.23	1.92	5.64	2.04	4.87	1.94
Forward Thinking	4.92	1.82	5.78	1.82	5.87	2.11	5.50	1.97	5.94	1.86
Detail Conscious	5.35	2.00	5.89	1.92	5.27	1.88	5.91	2.06	5.68	2.08
Conscientious	5.68	2.00	6.04	1.96	6.10	1.88	5.88	2.04	5.87	1.97
Rule Following	5.69	1.78	5.52	1.92	5.26	1.80	5.23	2.06	5.76	2.10
Relaxed	4.91	1.82	6.42	1.74	6.16	1.88	5.46	2.06	5.06	2.04
Worrying	7.28	1.51	4.47	1.83	5.03	1.99	5.44	2.05	4.74	1.93
Tough Minded	5.31	1.97	5.41	1.87	5.15	1.92	5.74	2.12	5.98	1.97
Optimistic	5.19	1.76	6.56	1.81	4.96	1.88	5.31	1.97	5.87	1.99
Trusting	5.60	1.87	6.55	1.80	4.16	1.85	5.30	1.82	5.33	1.87
Emotionally Controlled	5.52	1.92	5.08	1.93	4.85	2.07	5.61	1.97	5.36	1.86
Vigorous	5.13	1.94	6.53	1.78	4.68	2.06	6.36	1.78	6.06	1.84
Competitive	5.18	1.82	5.36	2.09	6.62	1.73	5.37	2.12	6.53	2.20
Achieving	5.04	1.97	5.68	1.84	6.01	1.99	5.87	2.00	6.44	1.86
Decisive	5.69	1.83	6.11	1.90	5.67	1.91	5.12	1.94	4.84	1.87

Table 44: Effect sizes by country or region.

Absolute Effect sizes are highlighted as follows:

Small: 0.2-0.5	Medium: 0.5-0.8	Large: >0.8

	Australia	Belgium (Dutch)	Belgium (French)	Brazil	Canada	China	Denmark	Finland	France	Germany, Austria and Switzerland
N	9,120	2,385	2,529	1,006	703	3,322	6,809	5,381	4,225	2,336
Persuasive	-0.02	-0.16	-0.27	0.17	0.01	-0.22	0.21	0.19	0.05	0.16
Controlling	-0.07	-0.52	-0.38	0.35	0.22	-0.25	0.21	-0.23	-0.03	0.18
Outspoken	-0.30	0.08	-0.05	-0.01	-0.27	-0.55	0.30	-0.18	-0.05	0.16
Independent Minded	-0.25	-0.09	-0.17	-0.21	-0.19	0.47	0.16	0.02	-0.17	0.34
Outgoing	-0.07	0.02	-0.18	-0.08	0.04	-0.09	0.52	0.37	-0.19	-0.07
Affiliative	0.02	0.33	0.12	-0.17	0.03	-0.37	0.42	-0.02	0.15	0.03
Socially Confident	0.15	-0.15	-0.03	-0.32	0.24	-0.15	0.41	0.20	0.00	0.26
Modest	0.16	-0.13	0.32	-0.17	0.14	0.15	-0.33	-0.20	0.28	-0.18
Democratic	0.11	0.44	0.18	-0.06	0.09	0.23	0.16	0.12	0.11	0.14
Caring	0.16	0.12	-0.30	-0.04	0.31	0.31	0.27	-0.04	-0.20	0.23
Data Rational	-0.04	-0.11	0.03	0.18	0.18	-0.15	-0.12	-0.40	0.11	-0.37
Evaluative	0.00	0.26	0.19	0.12	0.15	-0.36	0.24	-0.15	0.23	-0.01
Behavioural	0.10	0.04	-0.15	0.19	0.30	-0.16	0.28	-0.34	0.15	-0.10
Conventional	0.22	0.16	-0.04	-0.06	0.03	0.14	-0.16	0.13	-0.25	-0.60
Conceptual	0.21	0.09	0.14	-0.07	0.25	-0.03	0.00	-0.38	-0.01	0.07
Innovative	-0.04	-0.16	-0.32	-0.16	0.18	-0.12	0.02	-0.29	-0.26	0.22
Variety Seeking	-0.23	0.18	0.06	0.00	-0.04	-0.39	0.01	-0.22	0.30	0.32
Adaptable	-0.07	-0.27	0.37	-0.12	-0.14	0.59	-0.05	-0.08	0.53	0.13
Forward Thinking	0.02	-0.34	-0.15	-0.06	0.20	0.45	-0.08	-0.24	-0.11	-0.20
Detail Conscious	0.13	-0.22	0.20	-0.30	0.09	-0.01	0.16	-0.20	0.04	-0.08
Conscientious	0.08	0.08	0.14	-0.42	0.07	-0.09	0.23	-0.01	0.04	0.07
Rule Following	-0.01	-0.21	0.11	-0.24	-0.06	-0.99	-0.27	0.22	-0.01	-0.04
Relaxed	0.01	0.31	-0.17	-0.22	-0.17	-0.12	0.43	0.35	-0.20	0.15
Worrying	-0.07	-0.02	0.13	0.63	-0.11	0.04	-0.34	-0.10	0.04	-0.19
Tough Minded	0.10	-0.03	-0.34	-0.19	0.17	0.18	0.54	0.02	-0.40	0.12
Optimistic	0.01	-0.18	-0.53	0.03	0.01	0.05	0.46	0.32	-0.33	0.24
Trusting	-0.02	-0.10	-0.19	0.02	0.10	0.36	0.76	0.39	-0.10	-0.15
Emotionally Controlled	0.00	-0.18	0.31	-0.29	-0.06	0.02	-0.09	-0.18	0.51	0.01
Vigorous	0.32	0.28	-0.09	-0.39	0.26	-0.51	0.49	0.05	-0.17	-0.11
Competitive	-0.20	-0.50	-0.33	0.18	-0.04	0.32	0.10	-0.37	-0.14	-0.33
Achieving	0.02	-0.20	-0.40	-0.06	0.22	-0.47	0.19	-0.18	-0.11	0.23
Decisive	-0.28	-0.06	-0.11	0.30	-0.27	-0.44	0.16	0.39	0.08	0.03

	Greater China	Iceland	India	Italy	Korea	Latin America	Malaysia	Middle East	Netherlands	New Zealand
N	564	729	1,098	4,550	846	971	2,672	1,494	6,227	2,818
Persuasive	-0.27	-0.04	-0.06	-0.30	-0.11	0.46	-0.21	-0.23	-0.20	-0.17
Controlling	-0.54	-0.06	0.15	0.03	-0.63	0.56	-0.31	-0.09	-0.22	-0.11
Outspoken	-0.62	0.02	-0.07	0.26	-0.05	0.62	-0.24	0.10	0.24	-0.19
Independent Minded	0.21	0.29	0.25	0.20	-0.23	-0.10	-0.01	0.28	0.46	-0.33
Outgoing	0.12	-0.24	-0.30	0.15	0.17	-0.08	-0.01	-0.29	0.14	-0.12
Affiliative	-0.47	0.13	-0.35	-0.14	0.46	-0.32	-0.38	-0.43	0.56	-0.03
Socially Confident	-0.30	-0.04	-0.10	-0.28	-0.03	0.23	-0.17	-0.62	0.07	0.08
Modest	0.20	0.23	0.17	0.16	-0.35	-0.08	0.17	0.19	-0.13	0.14
Democratic	0.18	0.15	-0.32	-0.34	0.27	-0.62	-0.04	-0.20	0.15	0.12
Caring	0.11	0.09	-0.13	-0.29	-0.20	-0.12	-0.29	-0.27	0.34	0.03
Data Rational	-0.15	0.00	0.32	-0.17	-0.24	0.51	-0.06	0.02	-0.19	-0.01
Evaluative	-0.57	0.16	-0.21	0.15	-1.06	0.11	-0.56	-0.52	0.19	0.04
Behavioural	0.01	-0.29	-0.25	0.02	-0.23	-0.10	-0.06	-0.51	0.35	-0.06
Conventional	0.03	-0.10	0.14	-0.16	-0.01	-0.23	0.35	0.34	-0.29	0.31
Conceptual	0.30	-0.18	-0.13	-0.43	-0.29	-0.04	0.07	-0.07	0.19	0.20
Innovative	-0.39	0.08	0.13	-0.12	-0.46	0.39	-0.37	0.02	-0.04	-0.02
Variety Seeking	-0.26	0.47	-0.19	0.21	-0.28	0.08	-0.32	-0.21	0.40	-0.29
Adaptable	0.77	-0.46	0.09	0.11	0.61	-0.61	0.36	0.24	-0.11	-0.15
Forward Thinking	0.15	0.27	0.28	-0.38	-0.29	0.20	0.16	0.09	-0.34	-0.02
Detail Conscious	-0.27	0.10	-0.01	-0.44	-0.32	0.06	-0.07	0.19	-0.28	0.19
Conscientious	-0.32	0.01	-0.04	-0.17	-0.70	0.19	-0.38	0.05	-0.24	0.14
Rule Following	-0.70	0.23	0.29	-0.12	0.23	0.15	0.00	0.71	-0.44	-0.03
Relaxed	-0.34	0.30	-0.08	-0.22	-0.28	-0.32	-0.15	-0.25	0.55	0.17
Worrying	0.20	-0.30	-0.17	0.24	0.28	0.03	0.11	-0.12	-0.21	-0.07
Tough Minded	0.21	0.33	-0.26	-0.13	-0.23	0.04	-0.24	0.02	0.26	0.08
Optimistic	-0.30	0.31	0.25	-0.46	-0.21	-0.03	-0.10	-0.09	0.08	0.01
Trusting	-0.25	0.53	-0.06	-0.26	0.35	-0.12	-0.42	-0.40	0.20	0.04
Emotionally Controlled	0.18	-0.07	0.21	0.11	0.24	-0.24	0.20	0.41	-0.17	-0.01
Vigorous	-0.71	0.67	-0.05	-0.21	-0.59	0.10	-0.59	-0.47	0.26	0.41
Competitive	0.04	-0.38	0.30	0.11	-0.23	0.29	0.06	0.48	-0.29	-0.20
Achieving	-0.58	0.15	0.20	-0.21	-0.67	0.36	-0.21	0.05	-0.17	-0.02
Decisive	-0.36	0.39	-0.10	0.35	-0.44	0.48	-0.30	-0.36	0.25	-0.17

	Norway	Poland	Portugal	Singapore	South Africa	Spain	Sweden	Turkey	UK	US
N	4,659	927	2,248	4,053	4,880	695	13,244	1,107	22,612	4,114
Persuasive	0.10	-0.04	-0.22	-0.21	-0.06	0.27	0.13	0.11	0.08	0.37
Controlling	0.27	-0.02	-0.06	-0.15	0.02	-0.14	0.18	0.36	0.06	0.29
Outspoken	0.30	0.25	0.04	-0.26	-0.02	-0.19	-0.15	0.75	-0.14	-0.38
Independent Minded	-0.41	0.37	-0.23	-0.22	0.08	-0.08	-0.17	0.16	-0.21	-0.14
Outgoing	0.28	-0.30	-0.25	-0.10	-0.18	-0.04	0.33	0.30	-0.02	0.17
Affiliative	0.42	-0.34	-0.05	-0.20	-0.51	0.15	0.53	0.26	-0.08	0.01
Socially Confident	0.33	-0.30	-0.20	-0.03	-0.15	-0.49	0.47	-0.11	0.24	0.44
Modest	-0.16	0.03	-0.05	0.21	0.06	-0.15	-0.05	-0.43	0.09	0.04
Democratic	0.43	-0.39	0.17	0.13	-0.24	-0.39	0.12	-0.13	0.17	-0.27
Caring	0.14	-0.47	0.03	-0.02	-0.19	0.03	0.44	-0.29	0.15	0.08
Data Rational	-0.04	-0.26	0.11	0.05	0.13	-0.01	-0.26	0.66	0.03	0.05
Evaluative	0.10	0.12	0.29	-0.11	-0.10	0.15	0.05	0.61	0.21	-0.06
Behavioural	0.15	-0.12	0.11	0.20	-0.34	-0.05	0.28	0.26	0.12	0.10
Conventional	-0.04	0.01	-0.05	0.24	0.27	-0.01	0.00	-0.44	0.02	0.14
Conceptual	-0.26	-0.20	0.25	0.13	0.09	-0.10	-0.07	0.32	0.17	0.16
Innovative	0.07	0.24	-0.13	-0.27	0.00	0.03	0.25	0.61	0.10	0.18
Variety Seeking	0.17	0.22	0.08	-0.14	-0.14	0.04	0.04	0.22	-0.11	-0.18
Adaptable	-0.01	-0.09	-0.13	0.34	-0.37	-0.02	0.28	-0.63	0.07	-0.32
Forward Thinking	0.18	-0.19	-0.15	0.20	0.08	-0.29	0.14	0.19	0.00	0.22
Detail Conscious	0.08	0.03	0.21	0.06	0.18	-0.08	0.19	-0.12	0.20	0.09
Conscientious	0.28	-0.34	0.06	0.01	0.09	0.09	0.27	0.30	0.19	0.18
Rule Following	0.30	-0.02	0.12	-0.18	0.21	0.09	0.01	-0.12	-0.13	0.13
Relaxed	0.40	-0.34	-0.16	-0.16	-0.06	-0.30	0.46	0.33	-0.02	-0.22
Worrying	-0.33	0.21	0.61	-0.01	0.02	0.89	-0.52	-0.23	-0.03	-0.38
Tough Minded	-0.03	0.07	-0.23	0.01	-0.14	-0.09	-0.05	-0.18	0.12	0.24
Optimistic	0.40	-0.27	-0.35	-0.16	0.01	-0.15	0.53	-0.27	-0.09	0.19
Trusting	0.54	-0.35	-0.16	-0.24	-0.50	0.05	0.53	-0.67	-0.10	-0.08
Emotionally Controlled	-0.17	0.02	-0.19	0.09	0.13	0.01	-0.21	-0.32	0.06	-0.07
Vigorous	0.15	0.04	-0.06	-0.38	0.03	-0.18	0.52	-0.41	0.43	0.28
Competitive	0.13	-0.06	0.01	-0.12	0.08	-0.16	-0.07	0.56	-0.06	0.52
Achieving	0.25	-0.11	0.13	-0.24	0.17	-0.23	0.09	0.26	0.19	0.47
Decisive	0.40	0.21	-0.07	-0.39	-0.12	0.10	0.30	0.09	-0.19	-0.33

Country and region differences graphs

Below aggregate personality scores and effect sizes across countries are displayed graphically in Figures 16 to 47. Mean personality scores are provided on the left and effect sizes on the right.

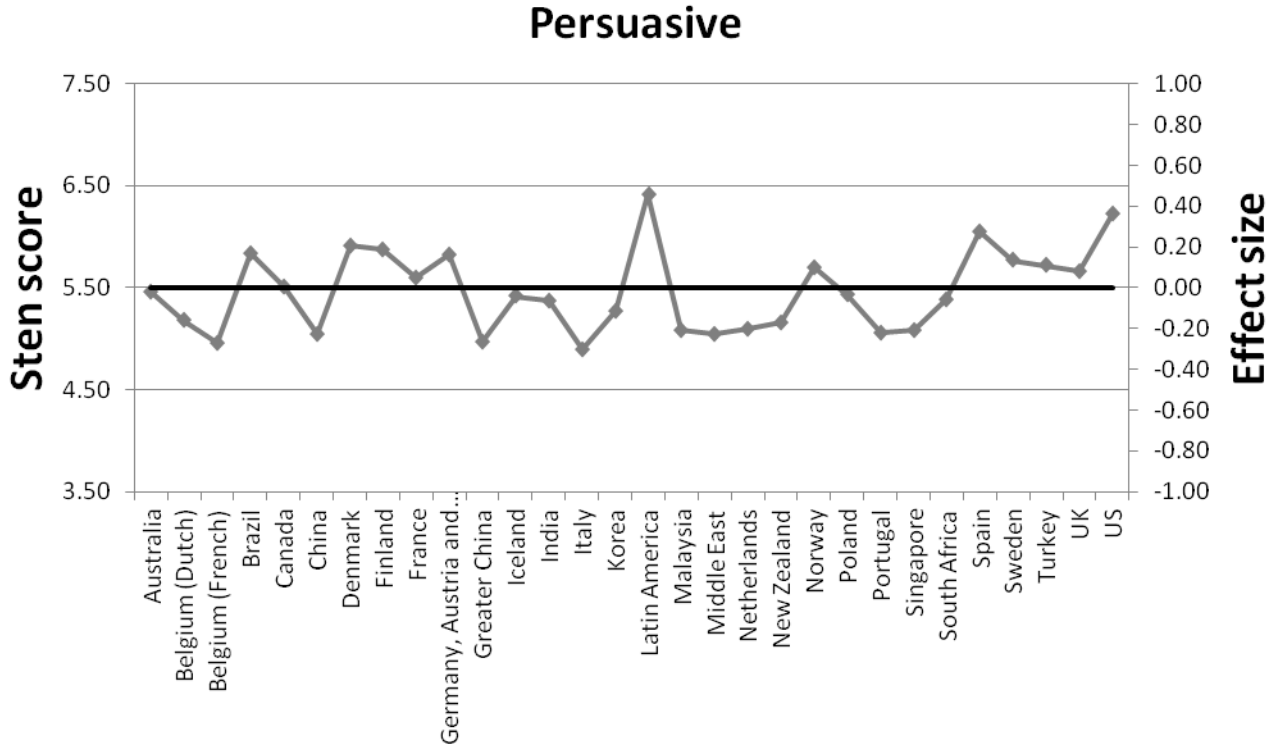**Figure 16: Country and region differences for Persuasive.**

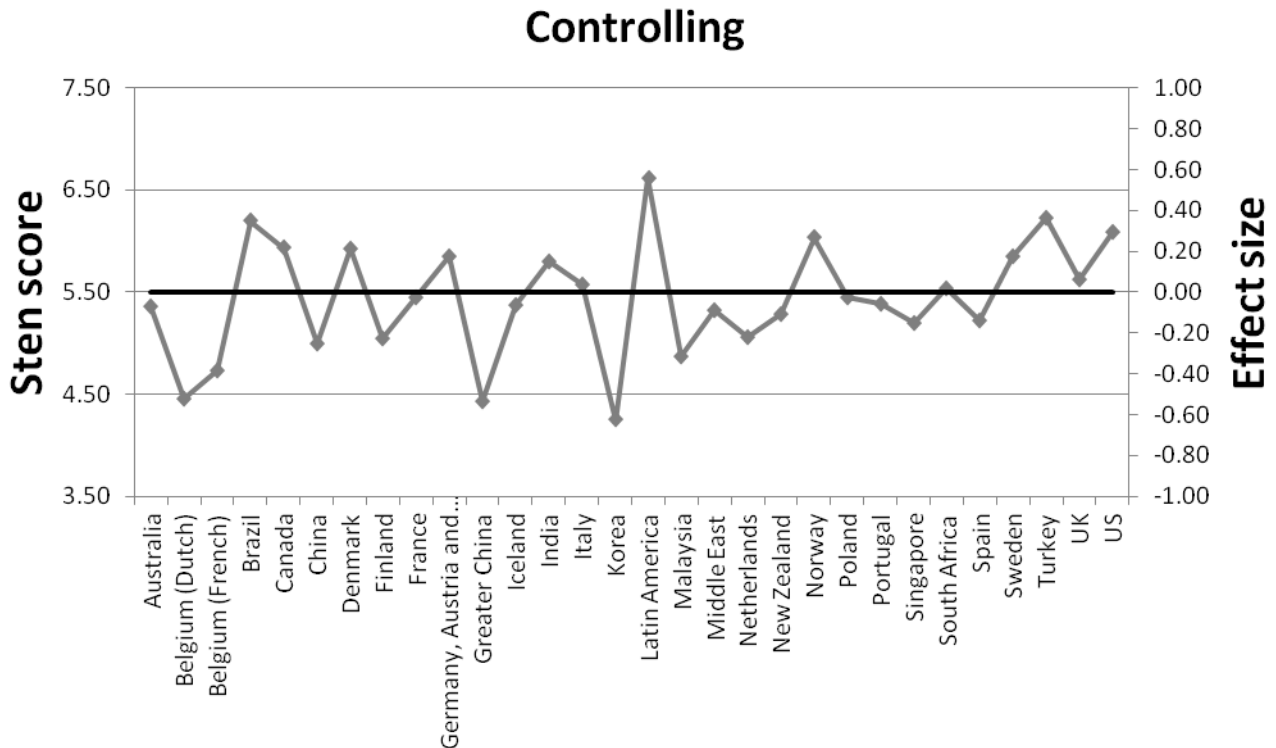**Figure 17: Country and region differences for Controlling.**

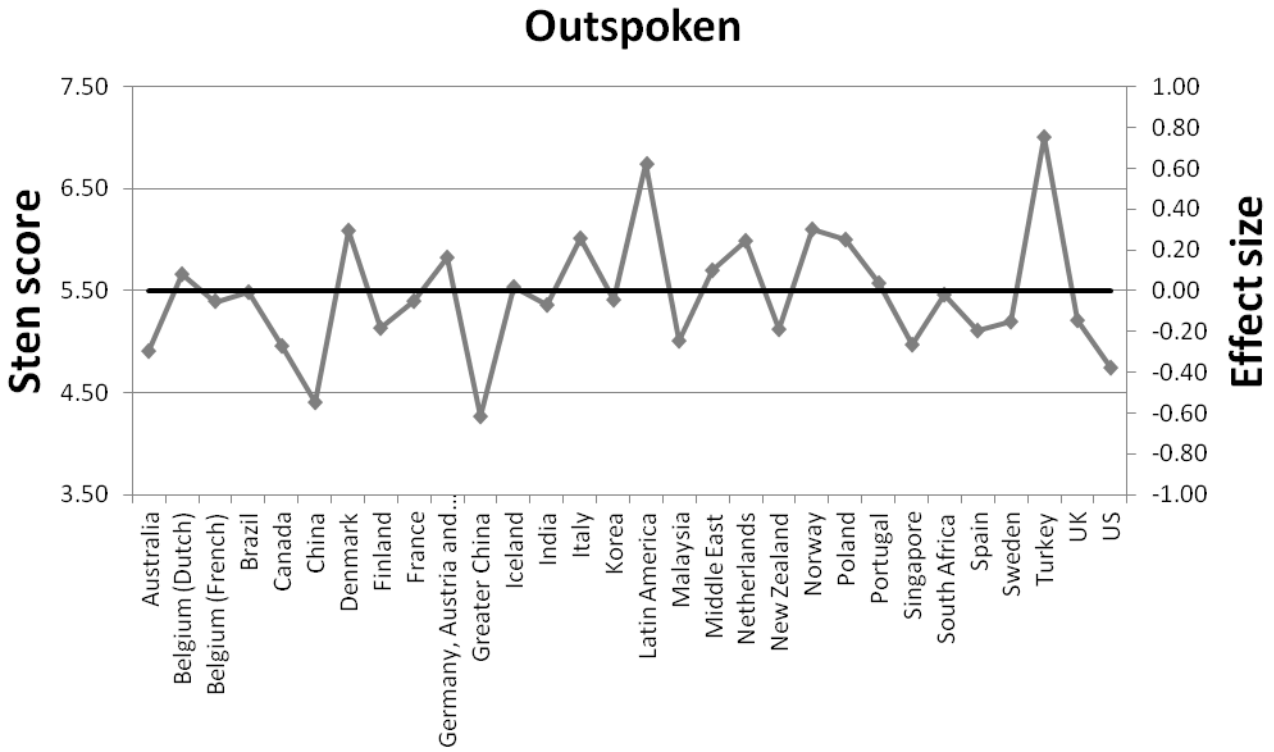**Figure 18: Country and region differences for Outspoken.**

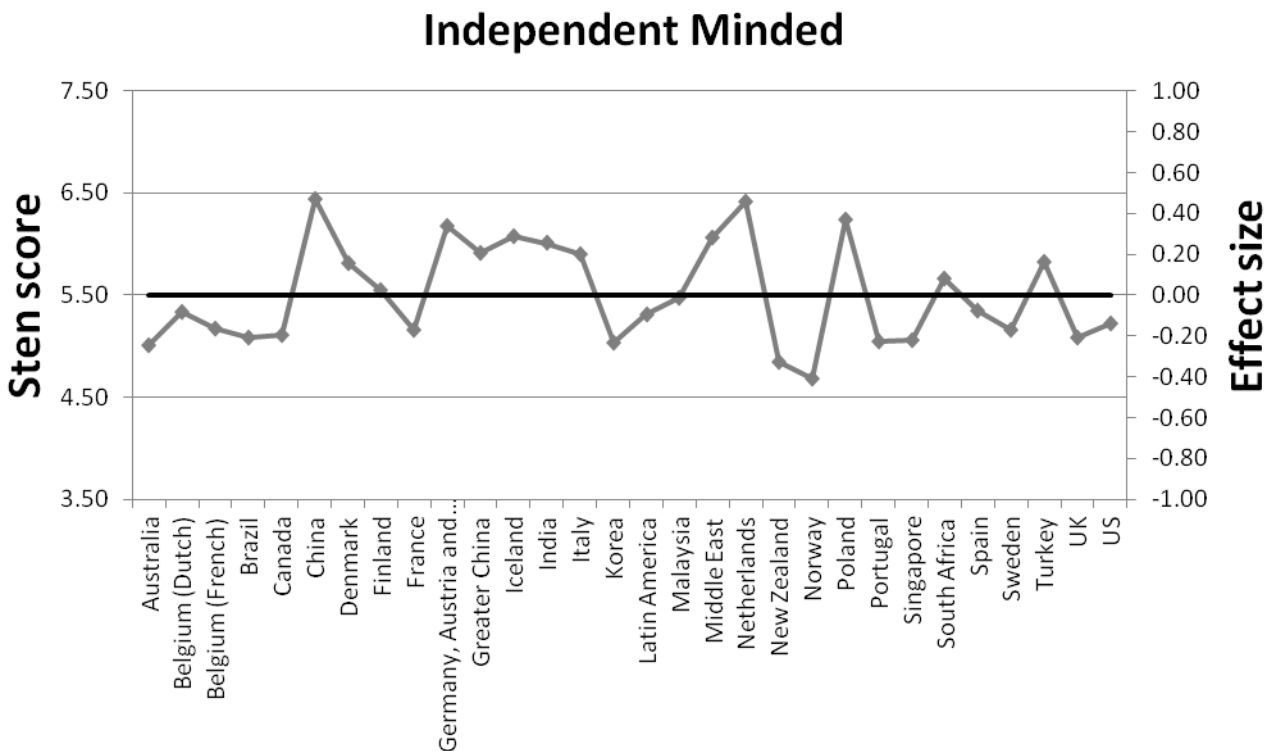**Figure 19: Country and region differences for Independent Minded.**

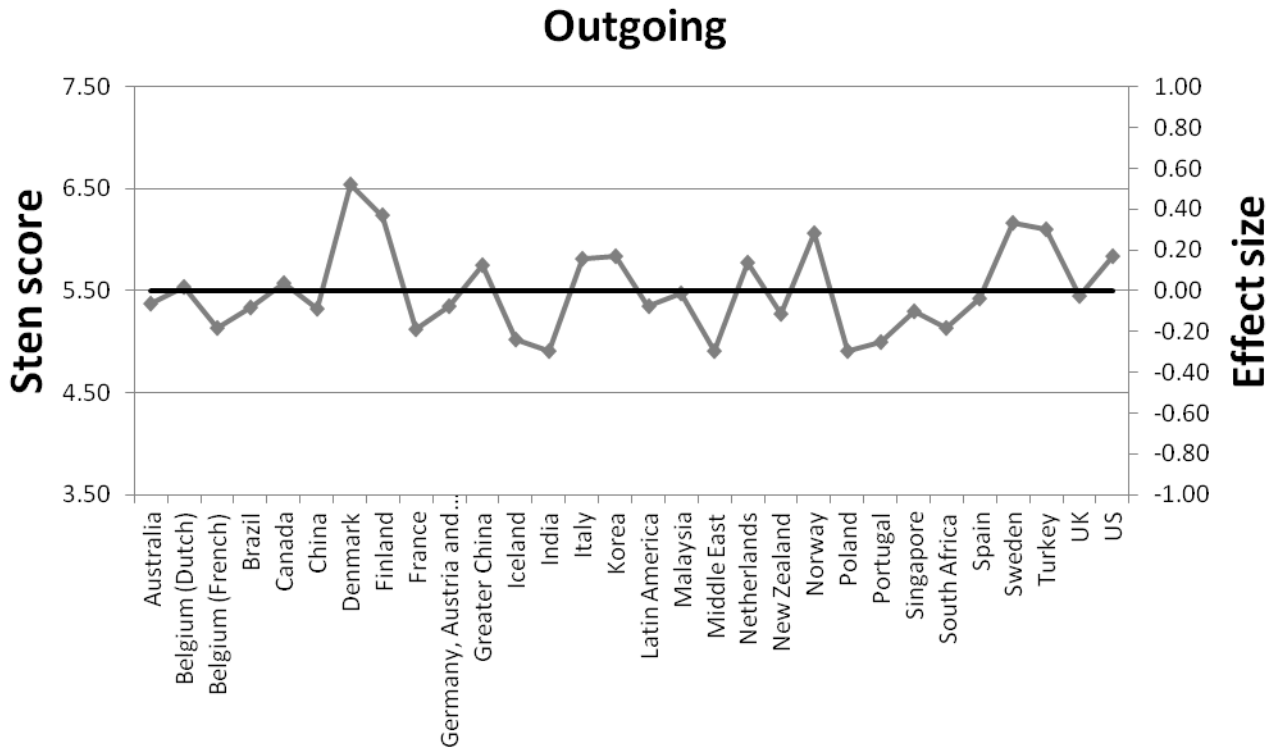**Figure 20: Country and region differences for Outgoing.**
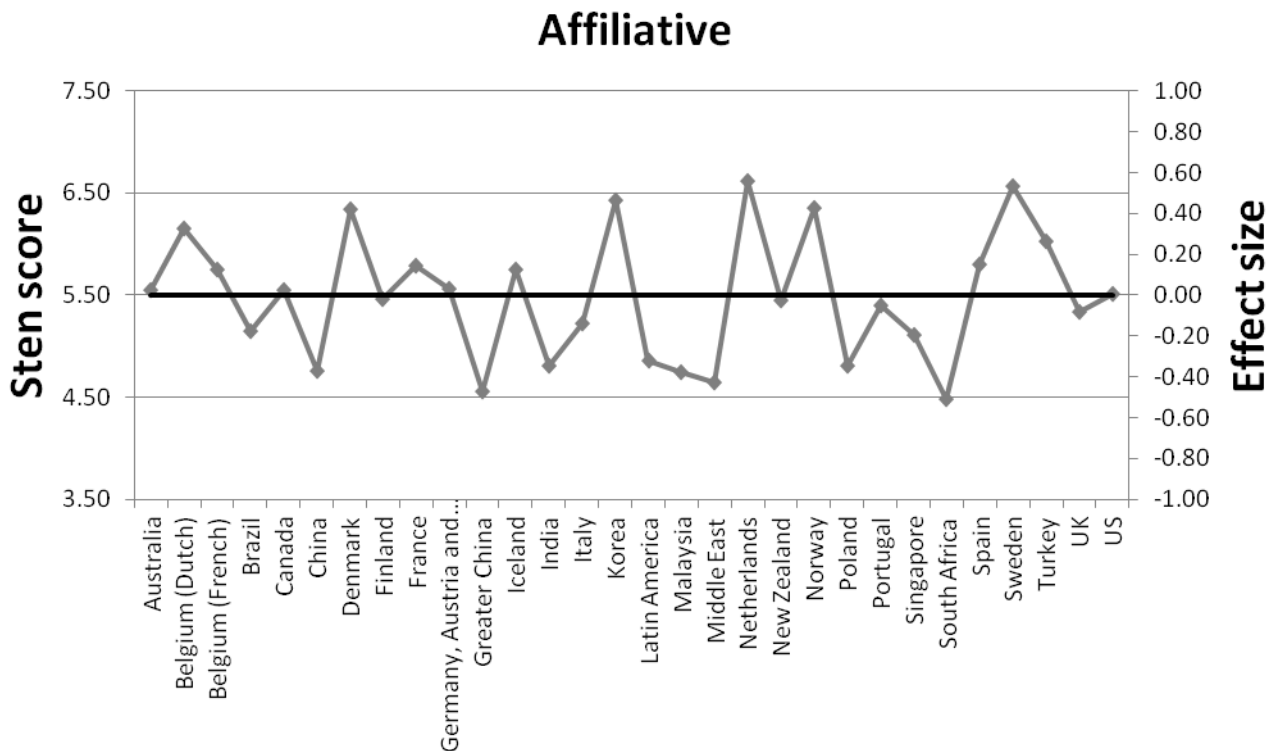

Figure 21: Country and region differences for Affiliative.

Figure 22: Country and region differences for Socially Confident.
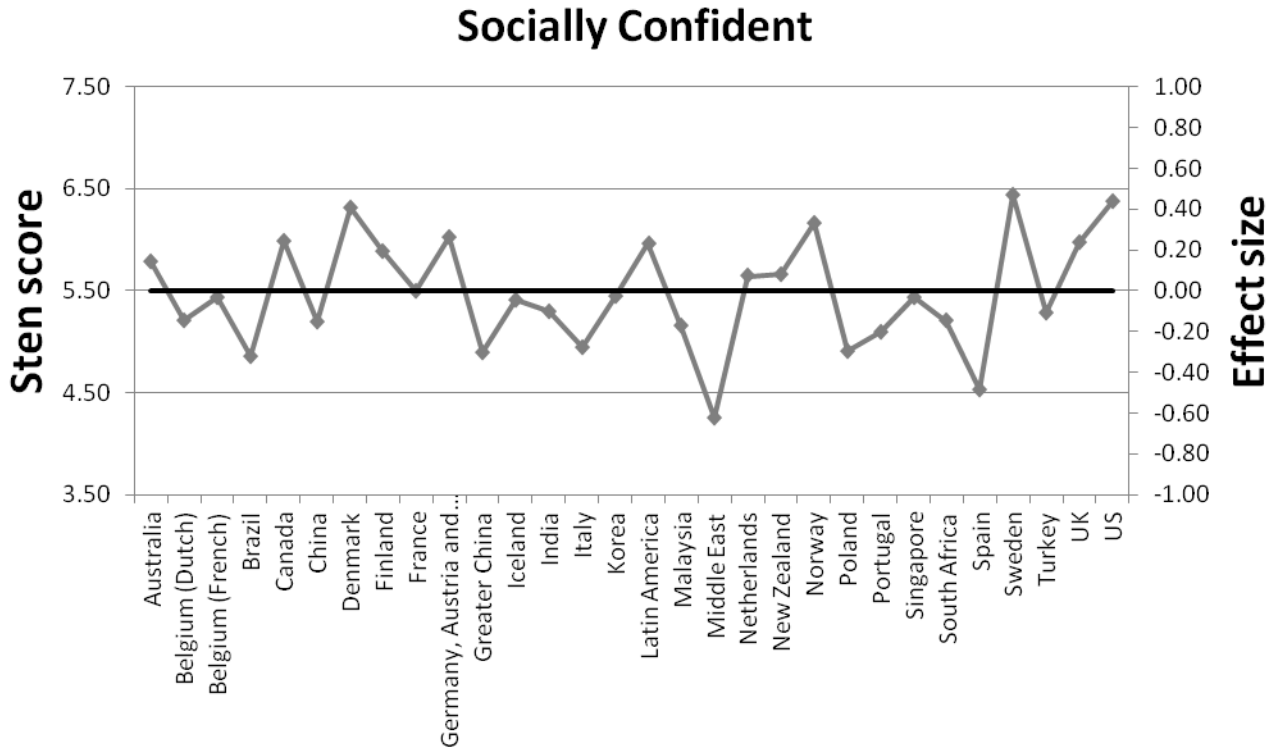

Socially Confident

Figure 23: Country and region differences for Modest.

Modest

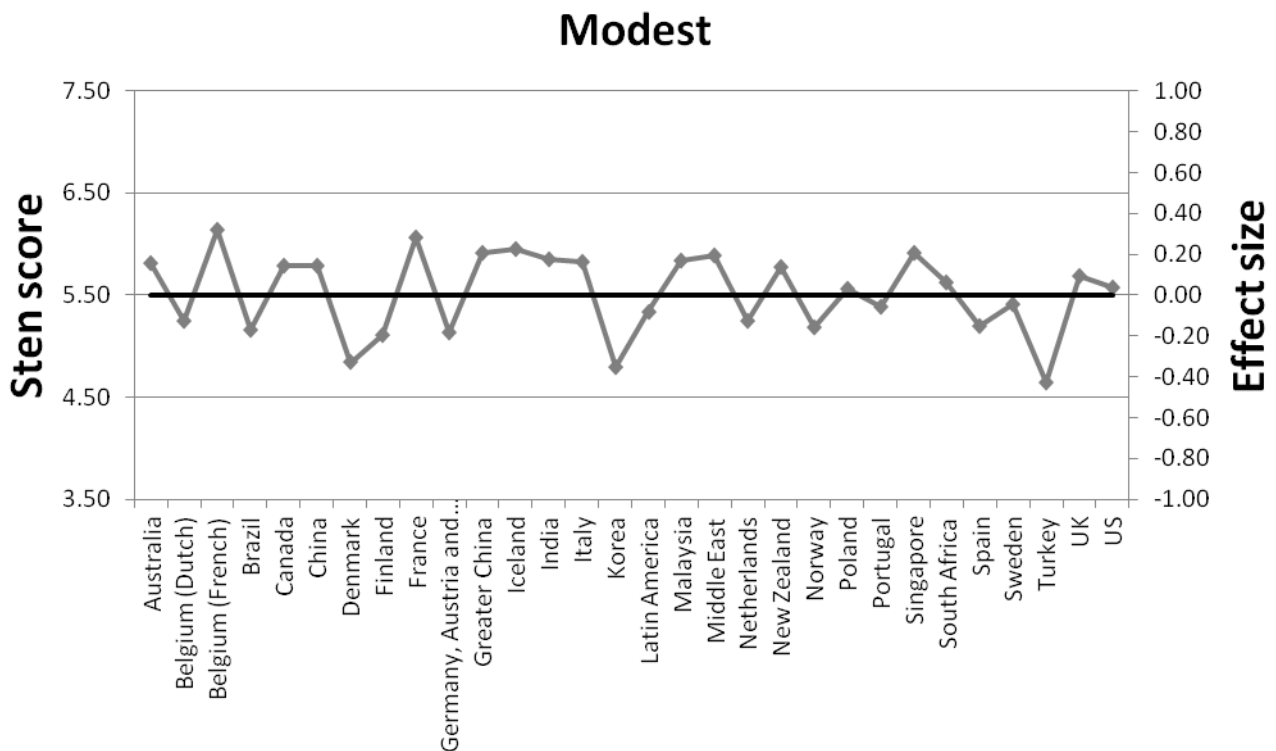**Figure 24: Country and region differences for Democratic.**
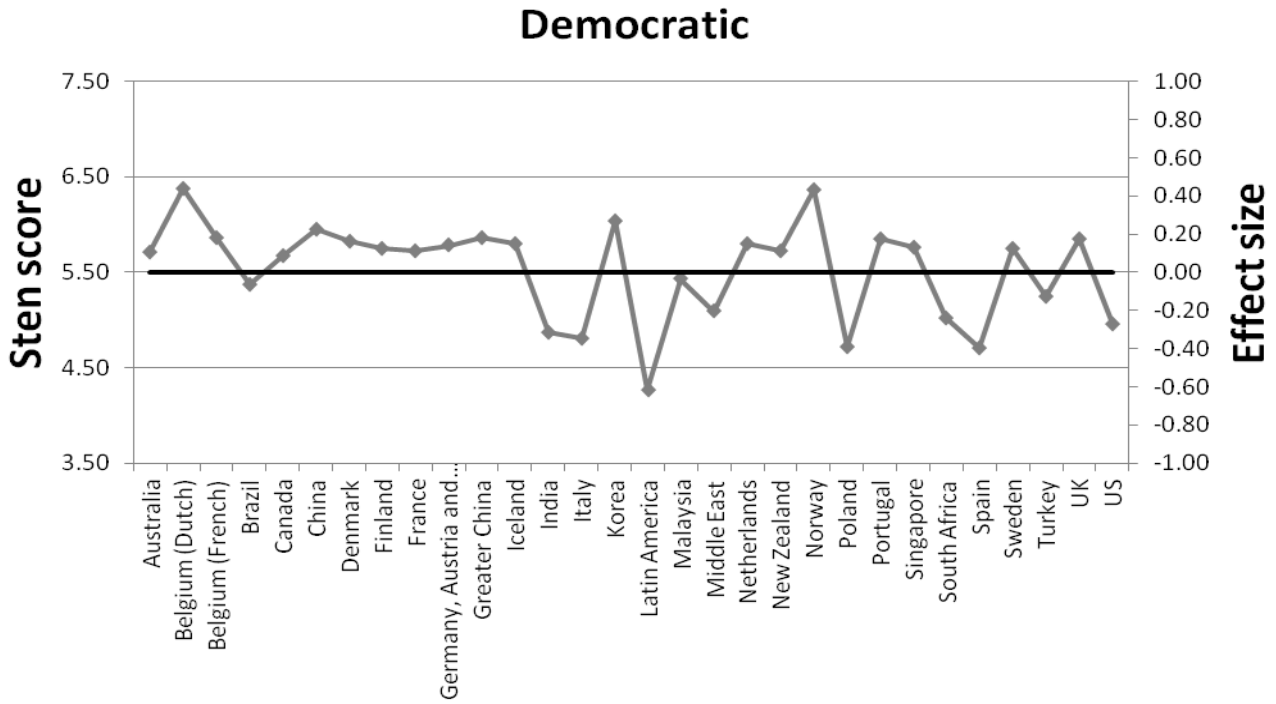
Figure 25: Country and region differences for Caring
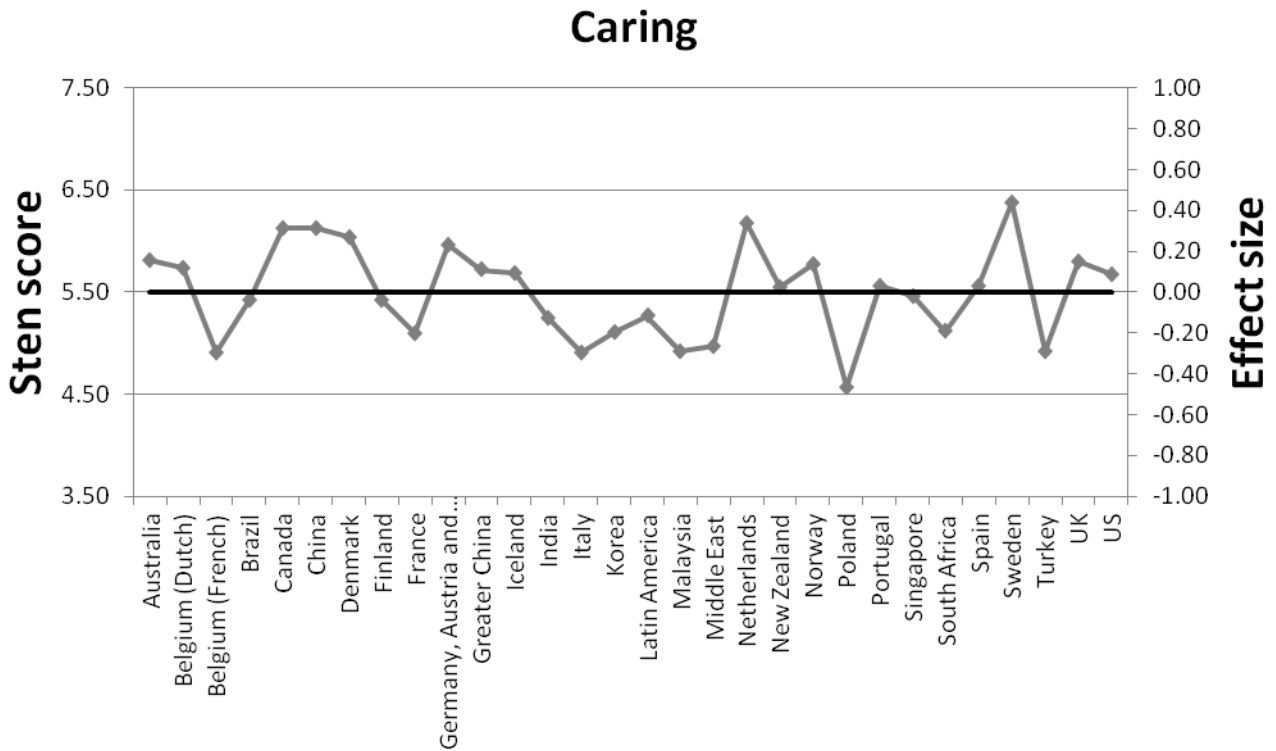
Figure 26: Country and region differences for Data Rational.

Data Rational
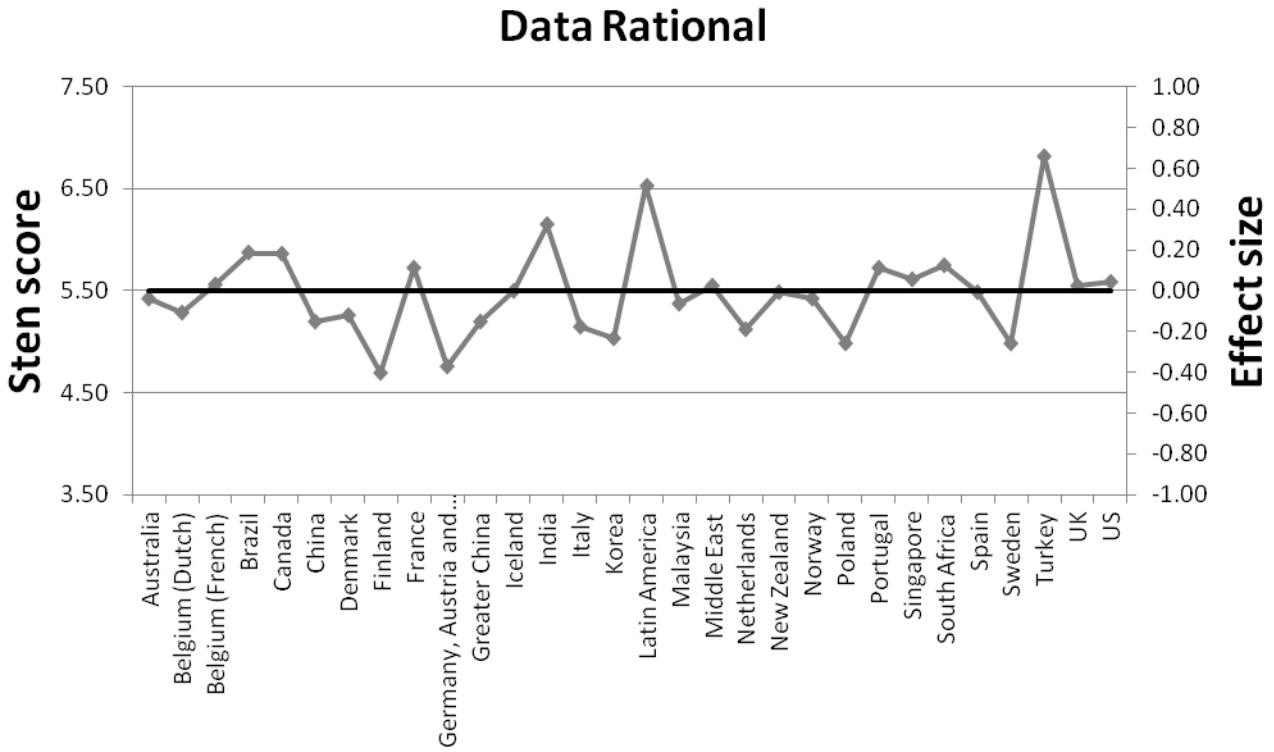
Figure 27: Country and region differences for Evaluative.

Evaluative
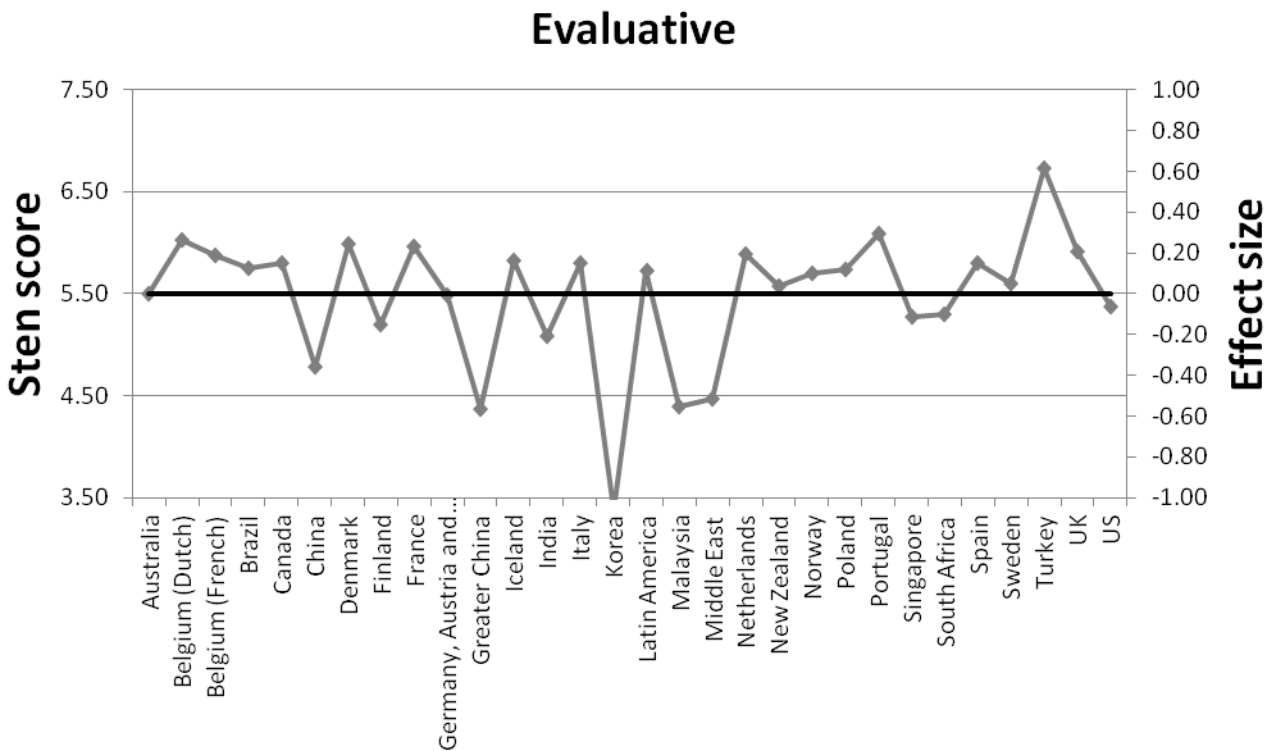
Figure 28: Country and region differences for Behavioural.
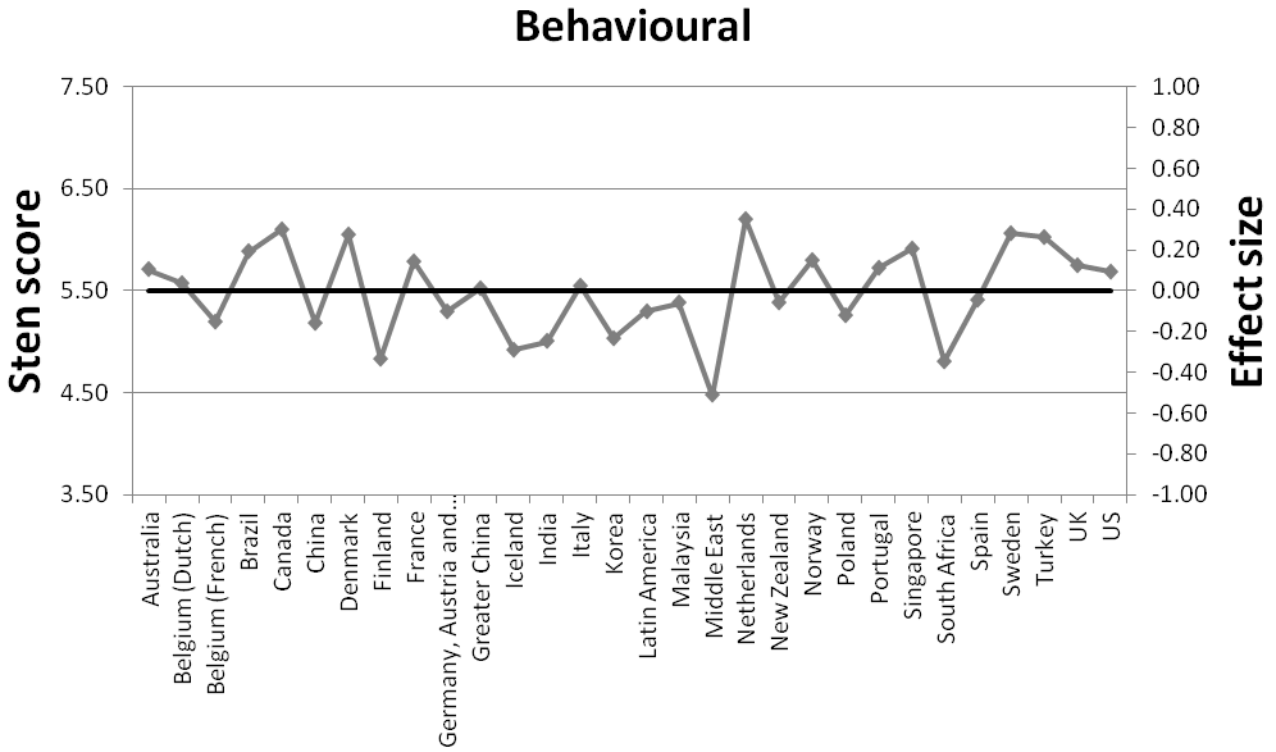
Figure 29: Country and region differences for Conventional.
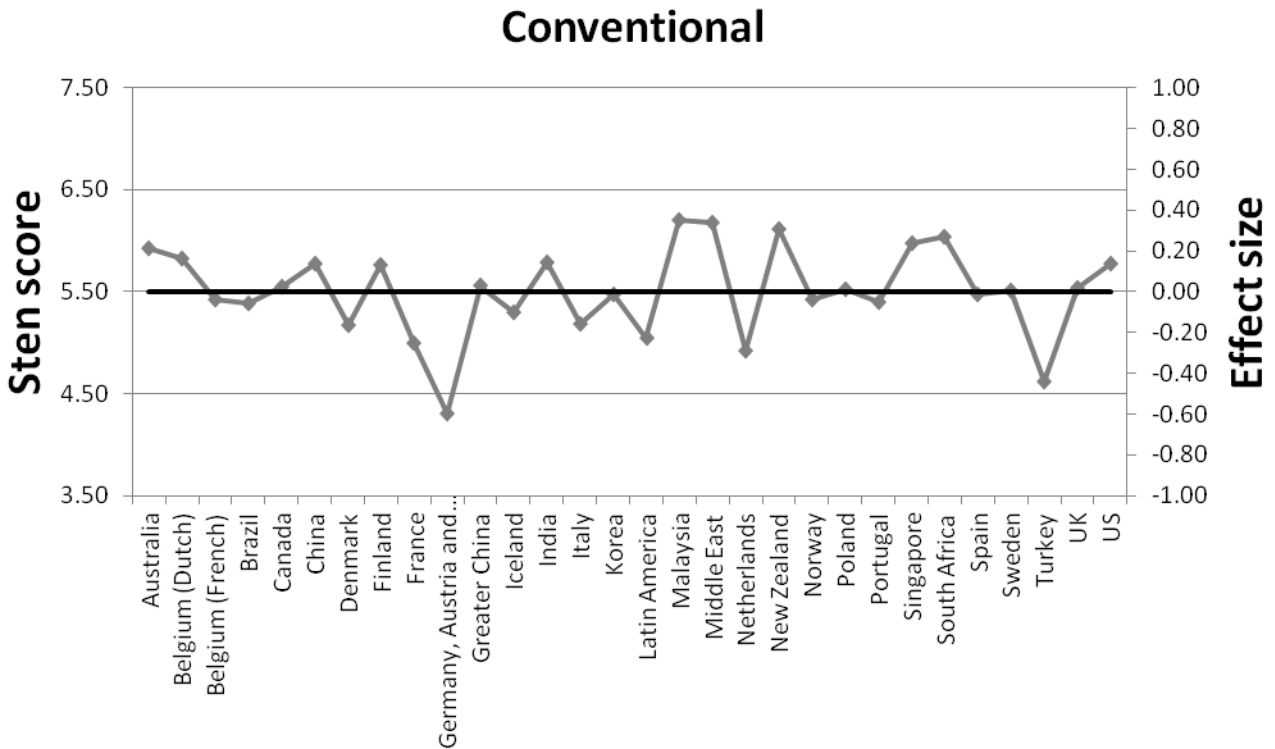
Figure 30: Country and region differences for Conceptual.

Conceptual
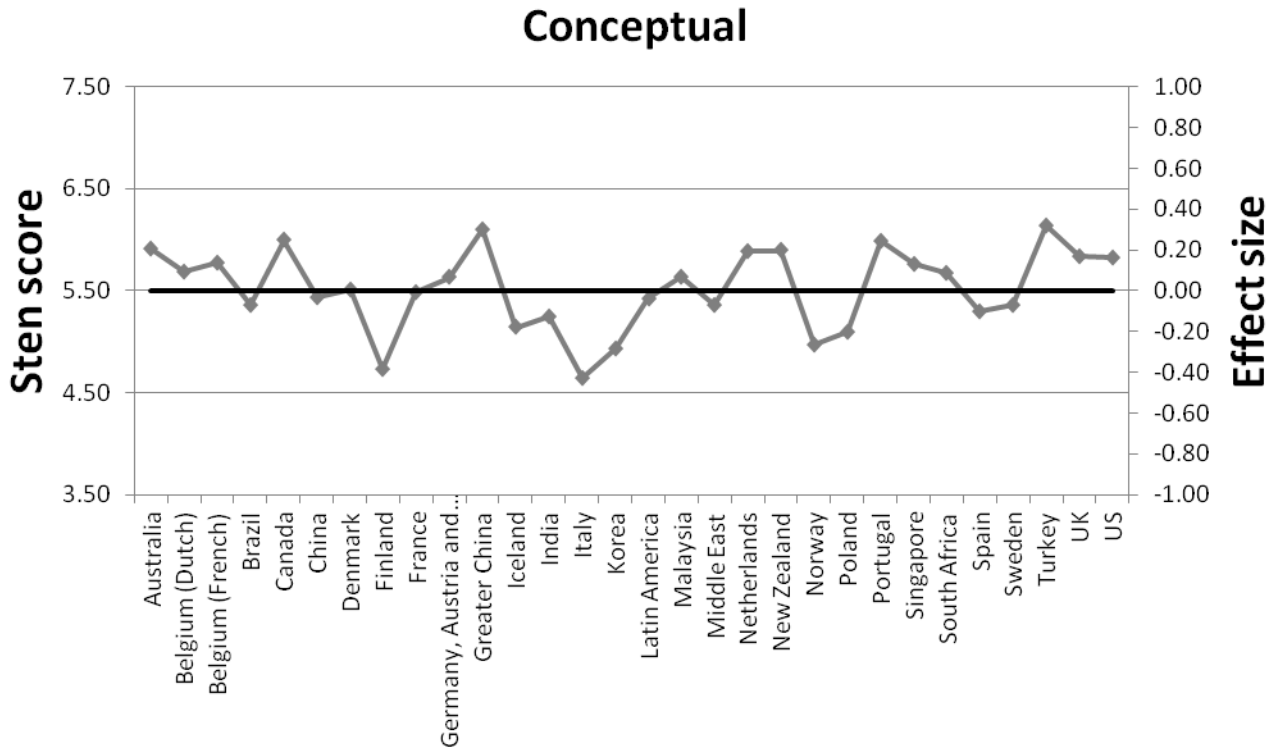
Figure 31: Country and region differences for Innovative.

Innovative
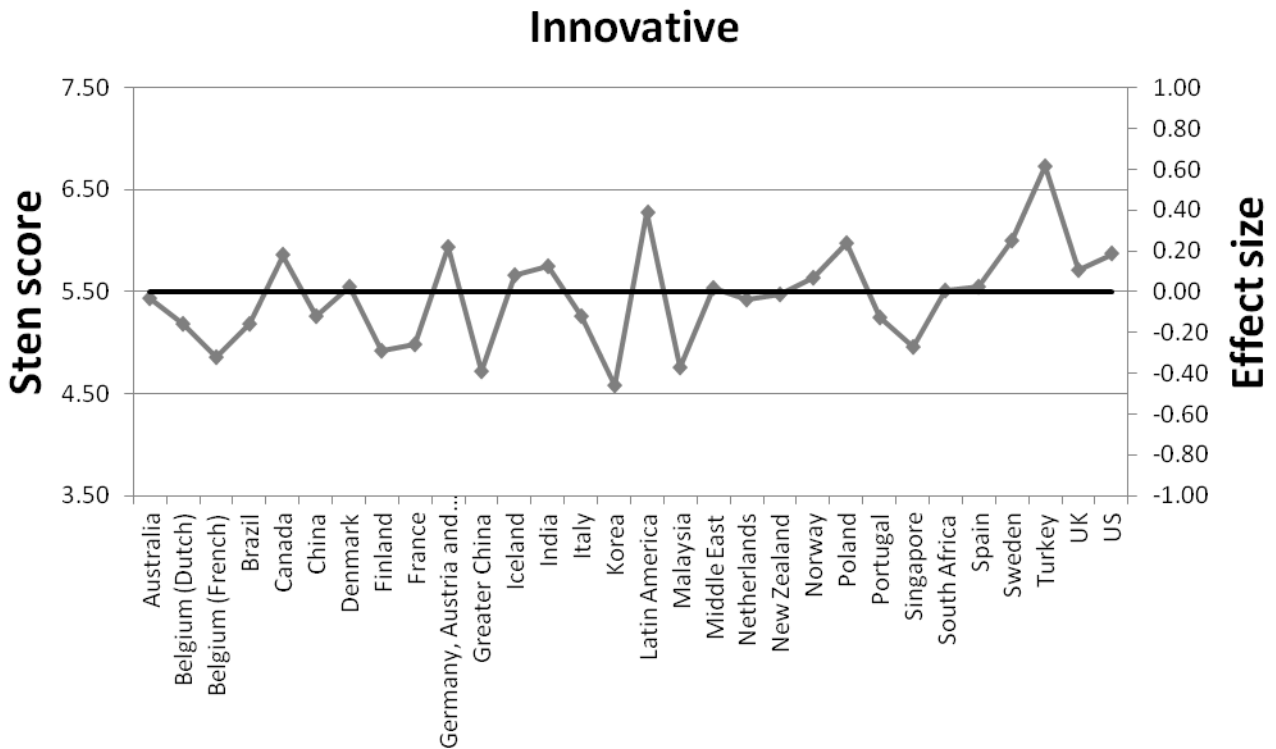
Figure 32: Country and region differences for Variety Seeking.

Variety Seeking
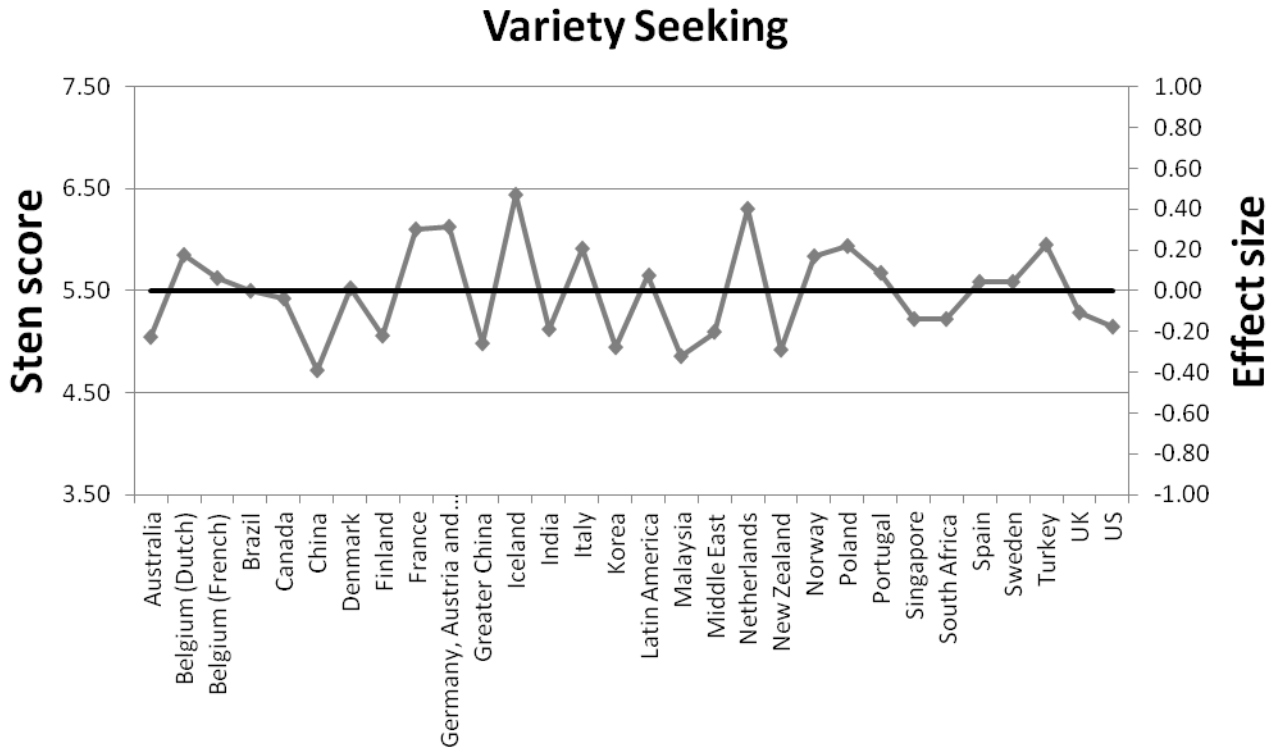
Figure 33: Country and region differences for Adaptable.

Adaptable
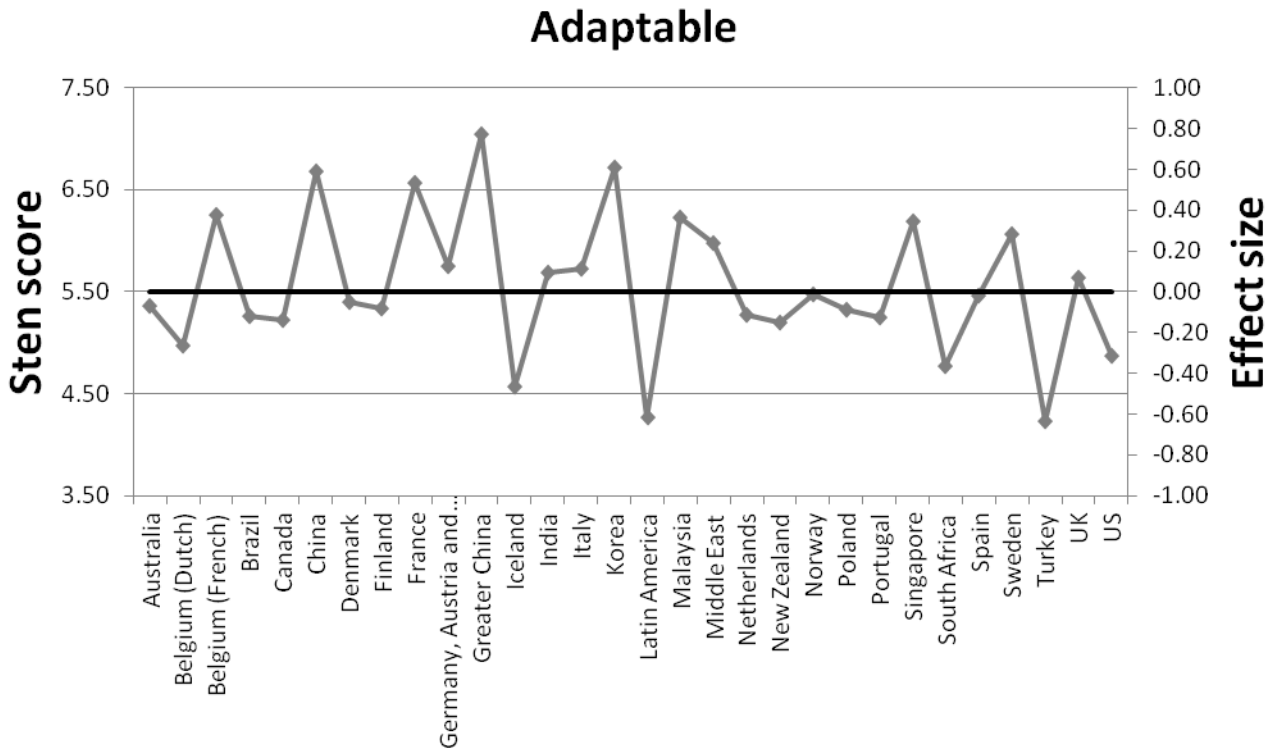
Figure 34: Country and region differences for Forward Thinking.

Forward Thinking
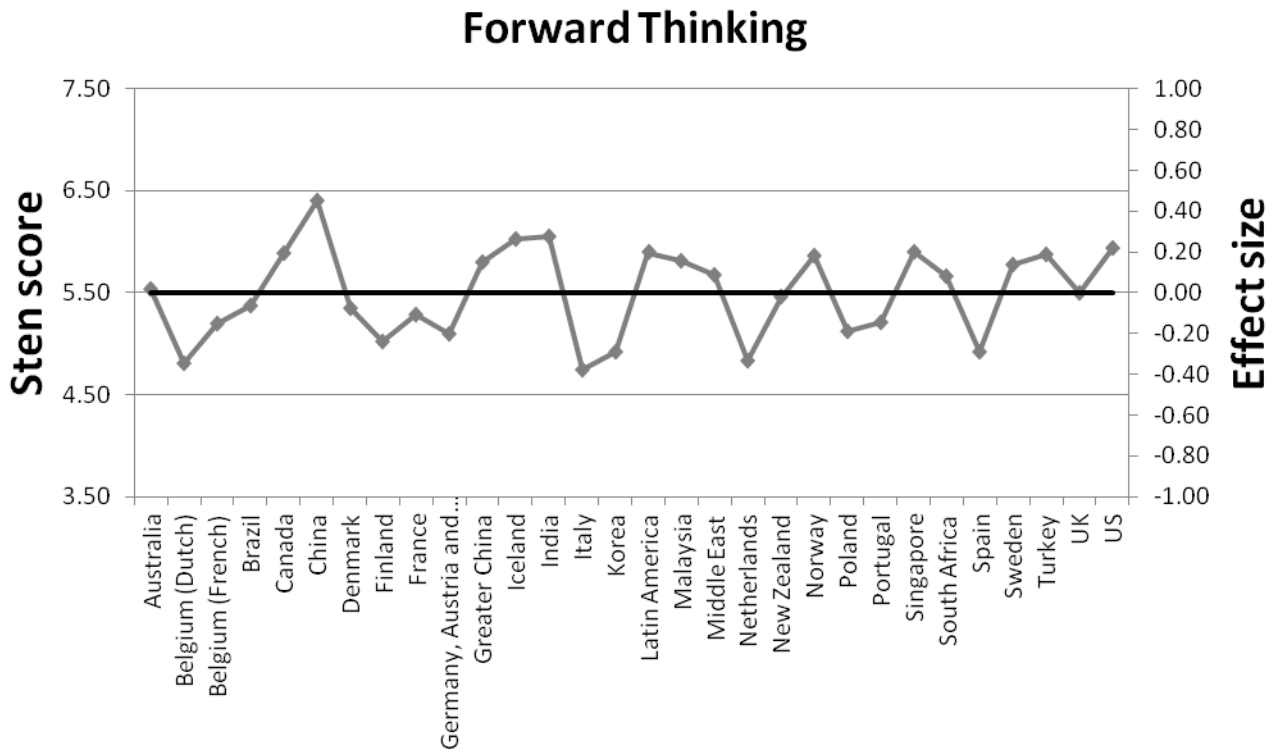
Figure 35: Country and region differences for Detail conscious.

Detail Conscious
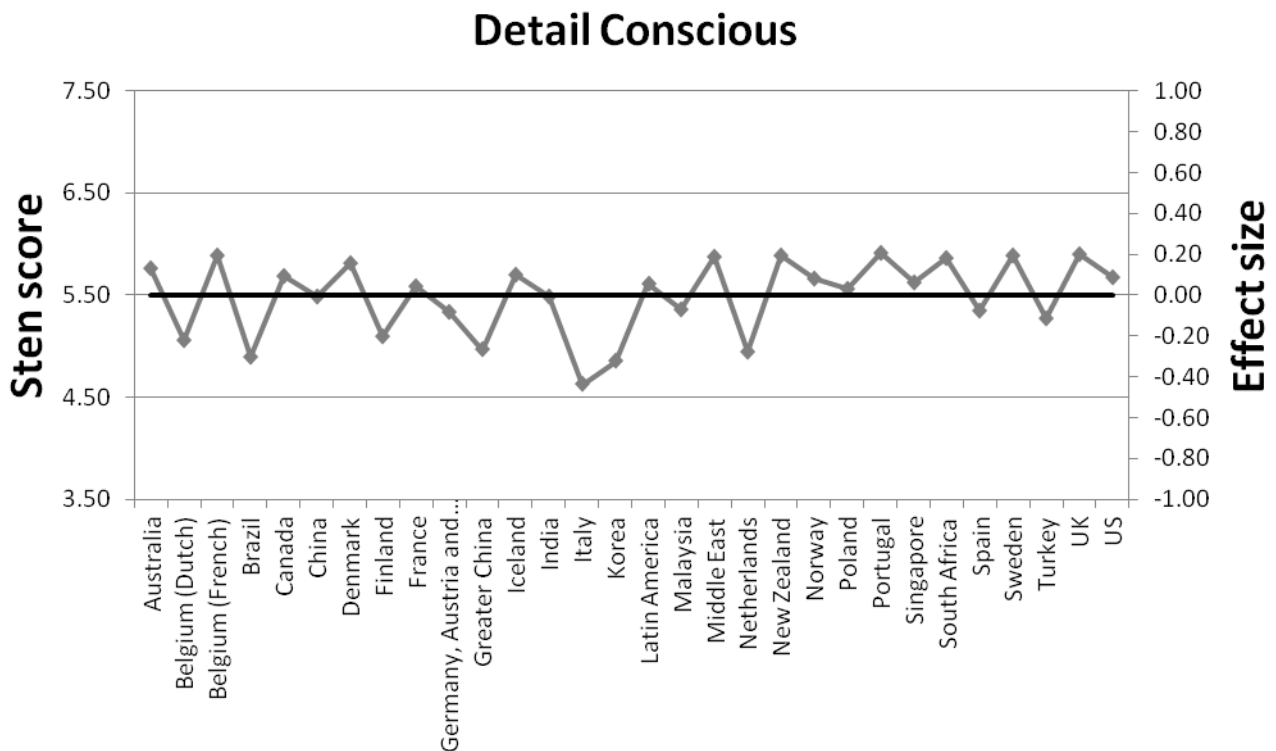
Figure 36: Country and region differences for Conscientious.
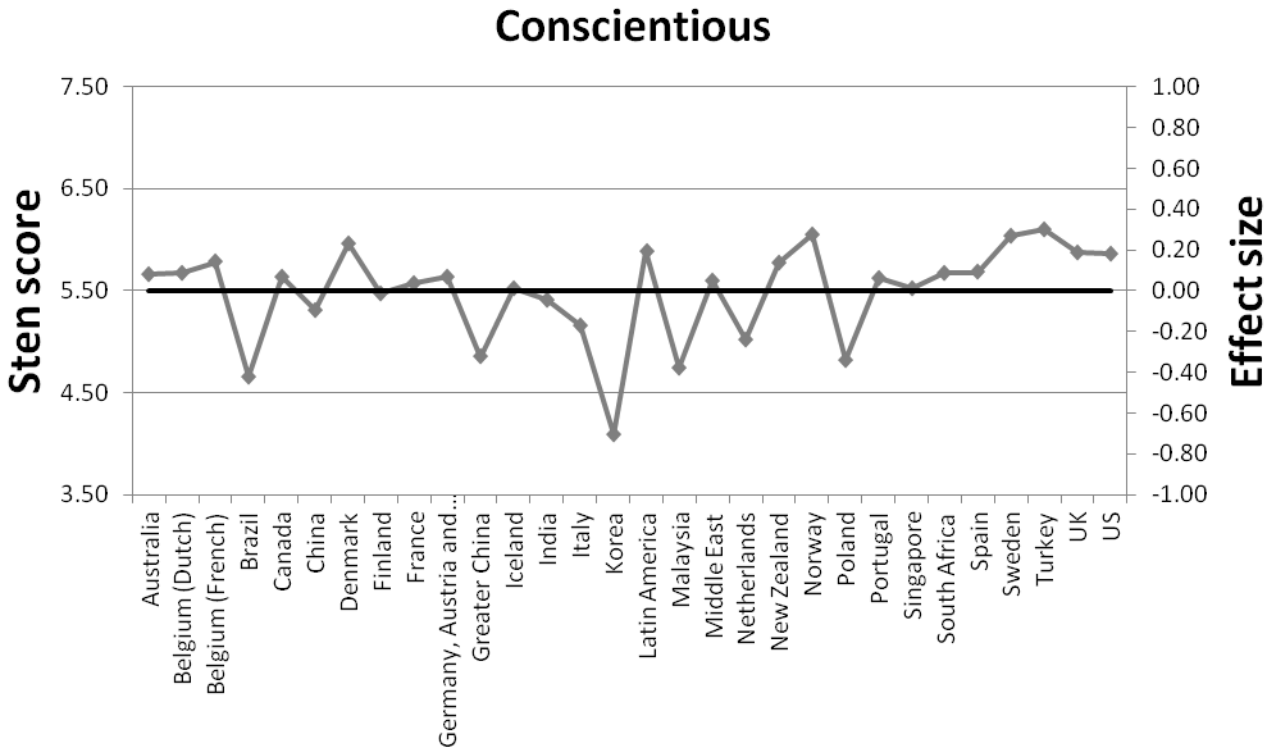
Figure 37: Country and region differences for Rule Following.
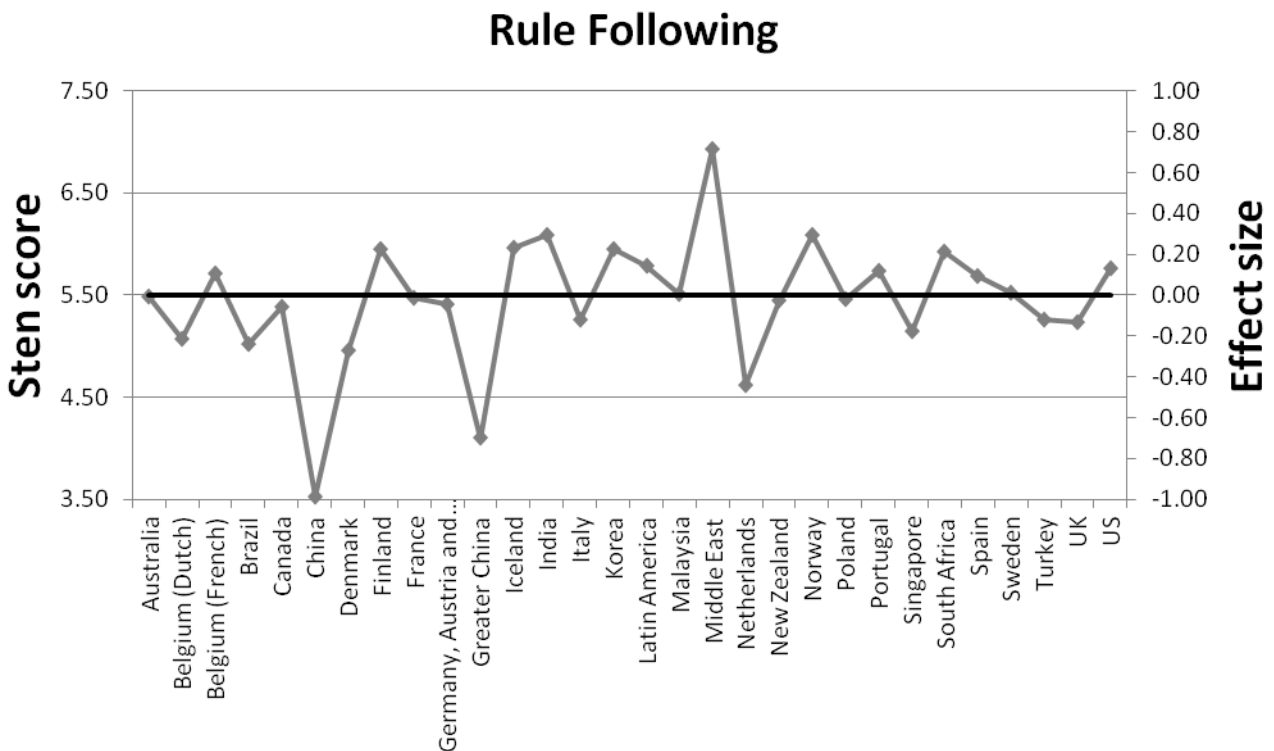
Figure 38: Country and region differences for Relaxed.
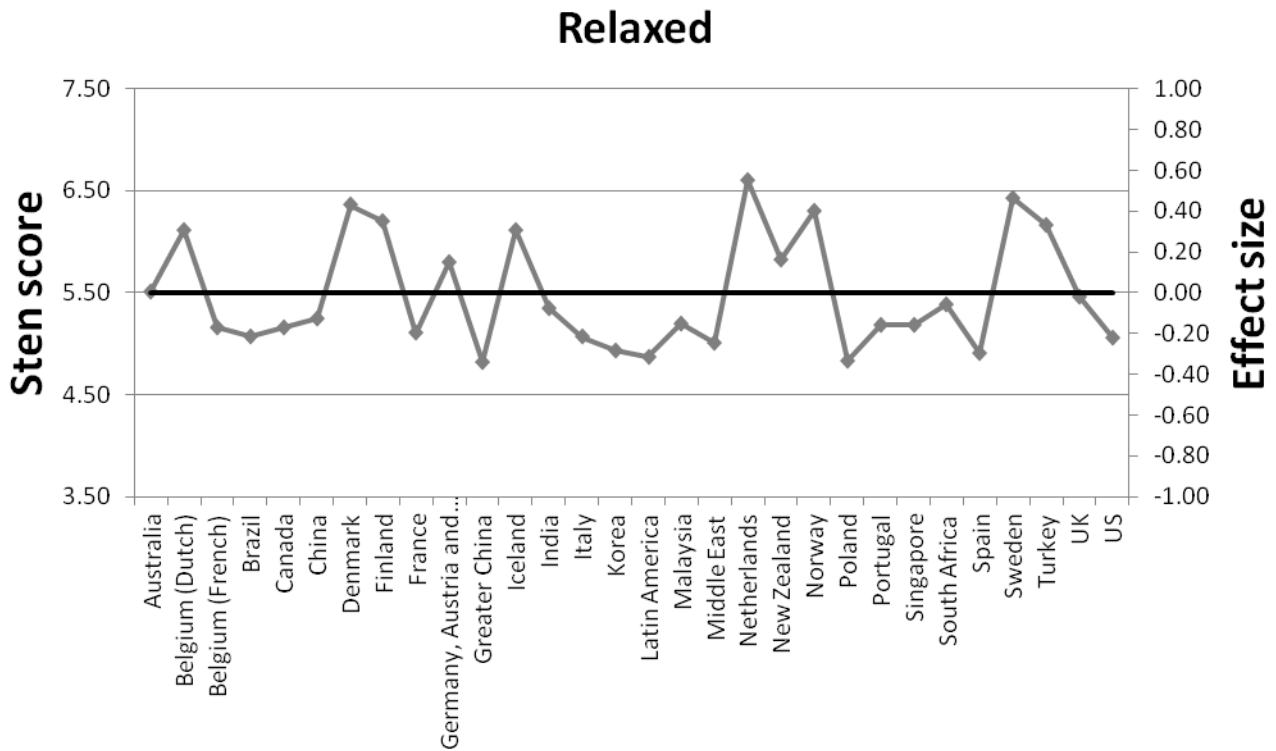
Figure 39: Country and region differences for Worrying.

Figure 40: Country and region differences for Tough Minded.

Figure 41: Country and region differences for Optimistic.
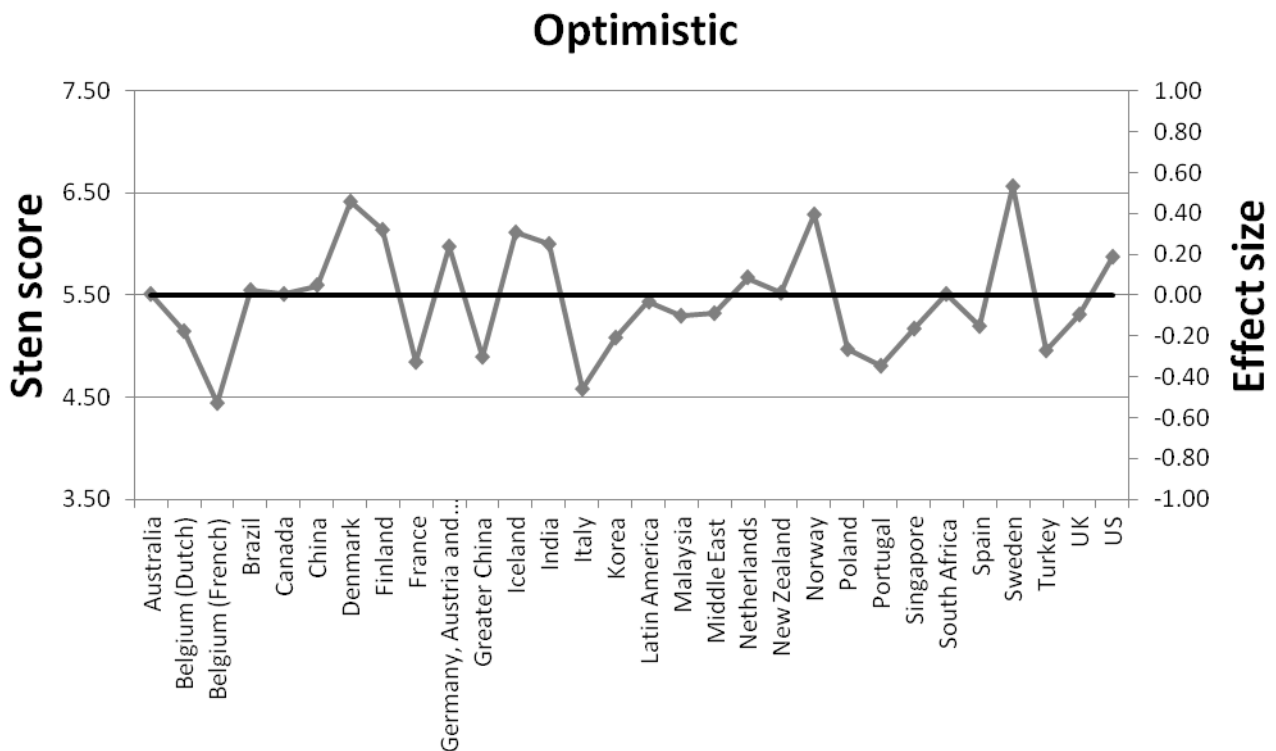
Figure 42: Country and region differences for Trusting.

Trusting
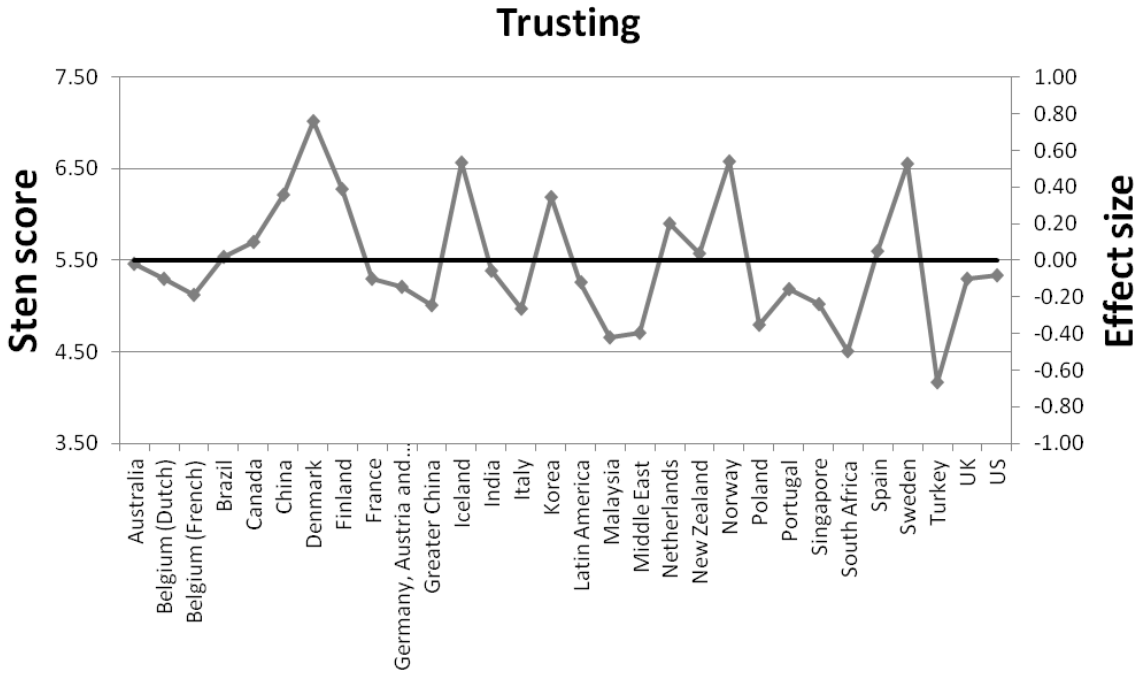
Figure 43: Country and region differences for Emotionally Controlled.

Emotionally Controlled
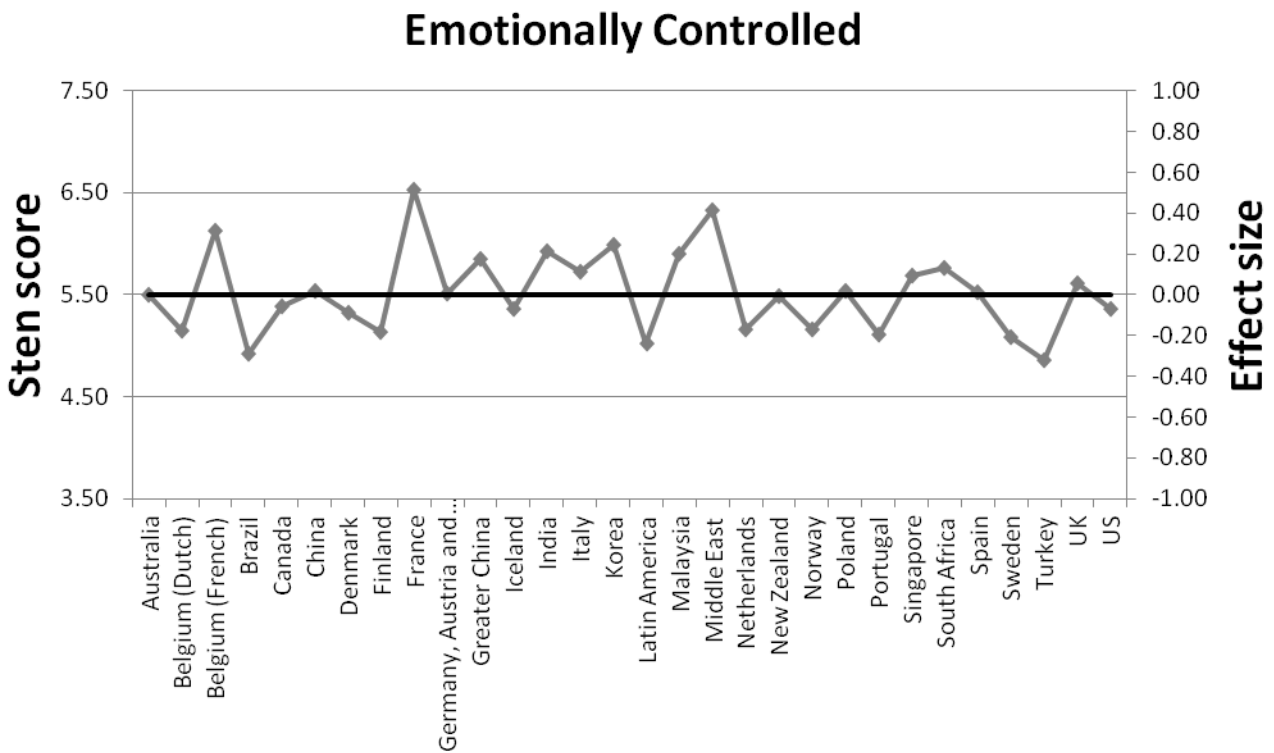
Figure 44: Country and region differences for Vigorous.
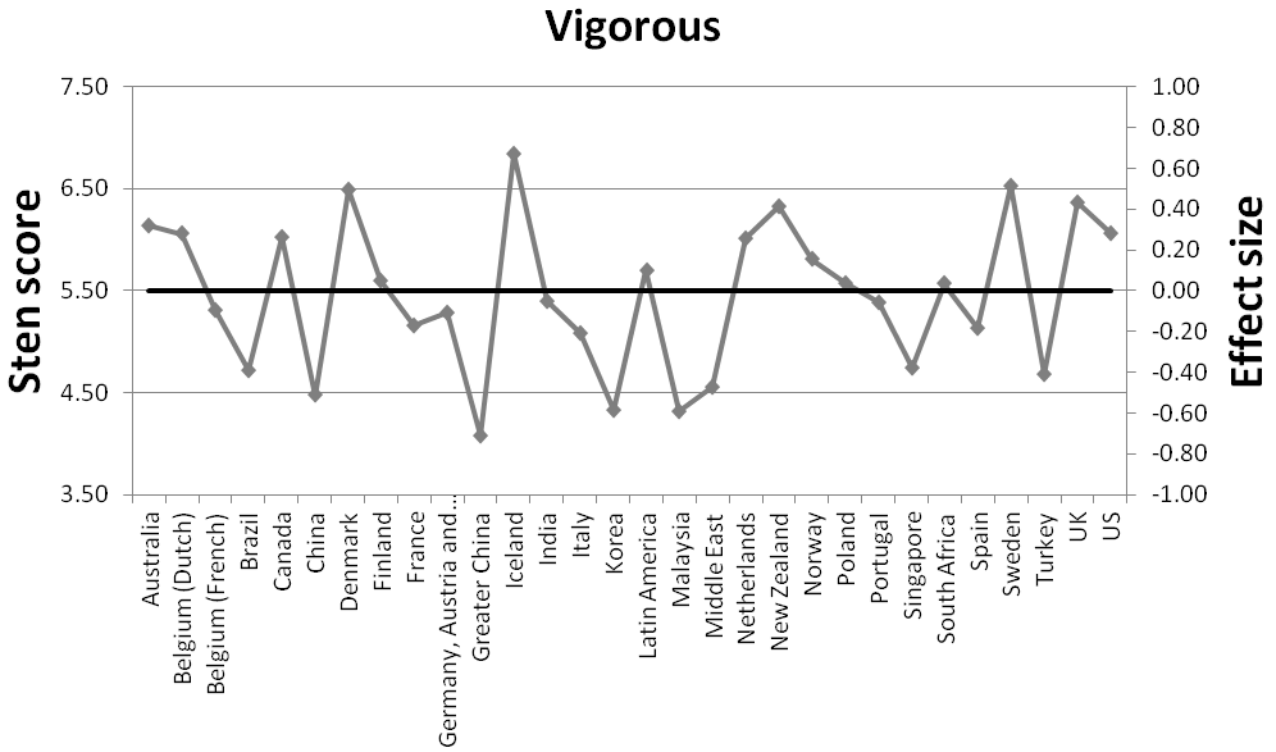
Figure 45: Country and region differences for Competitive.
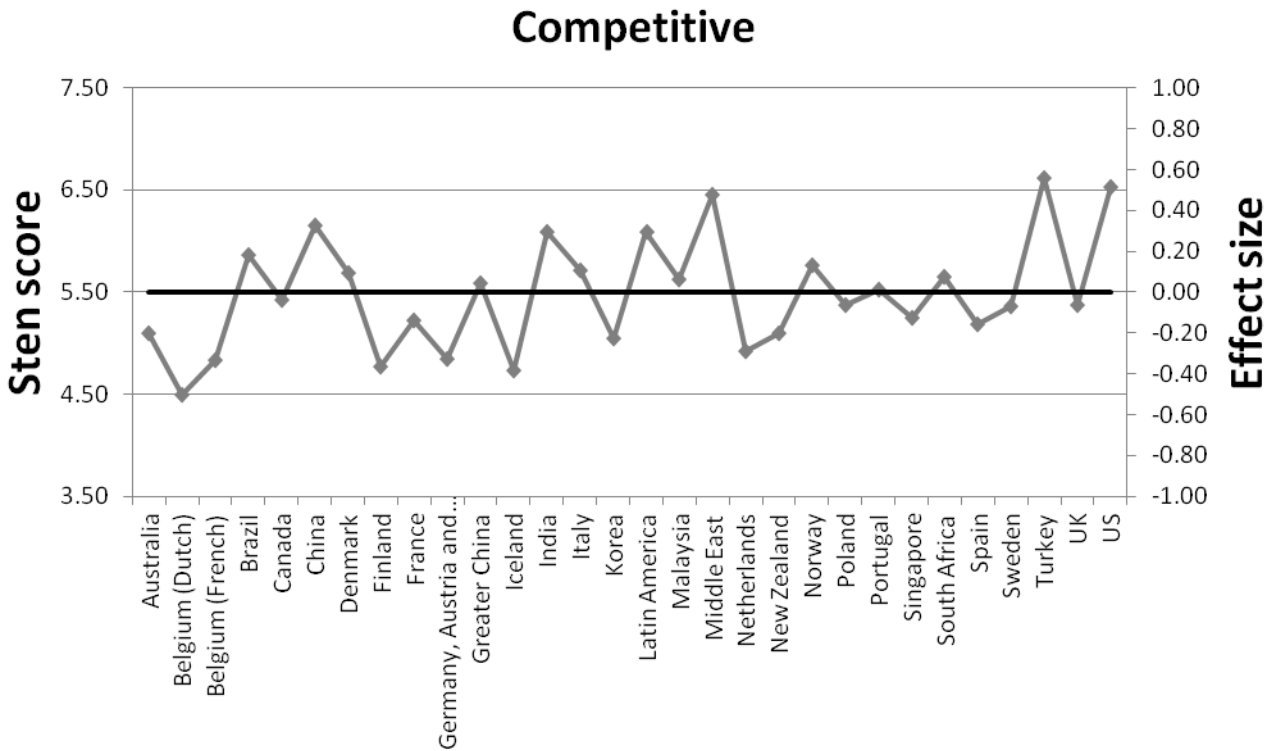
Figure 46: Country and region differences for Achieving.
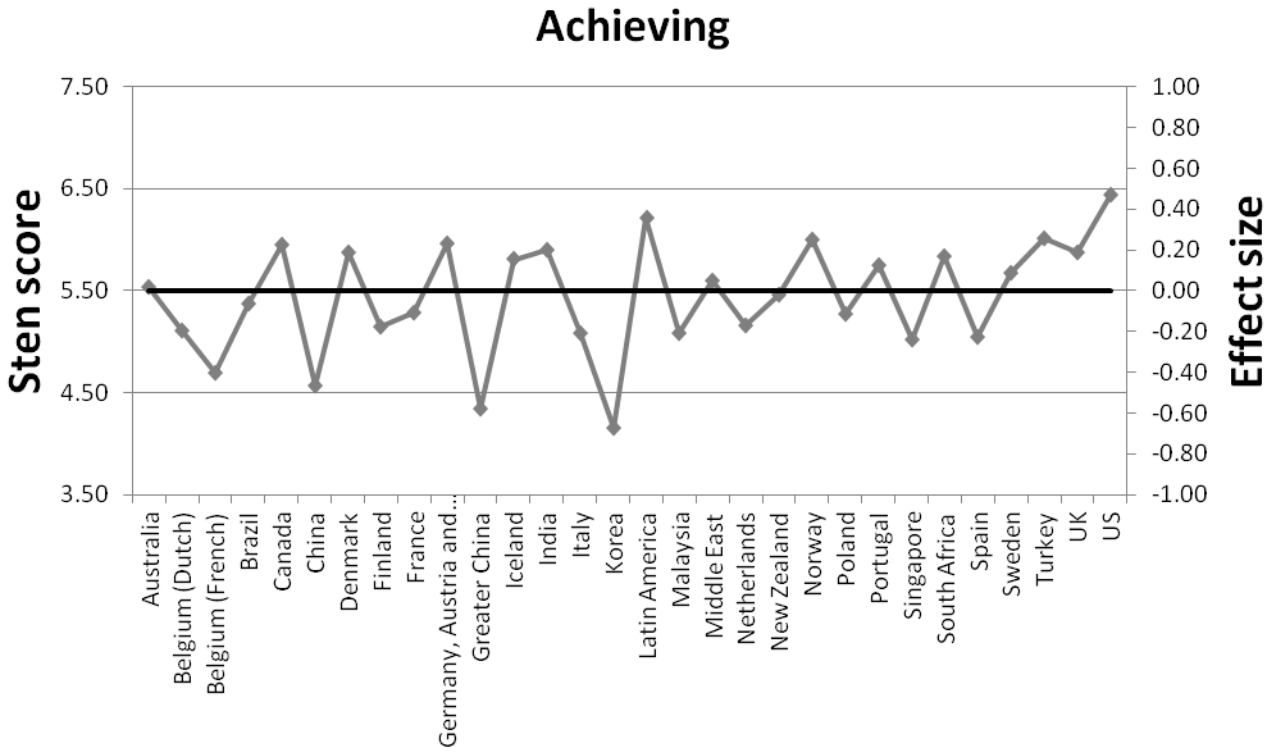
Figure 47: Country and region differences for Decisive.
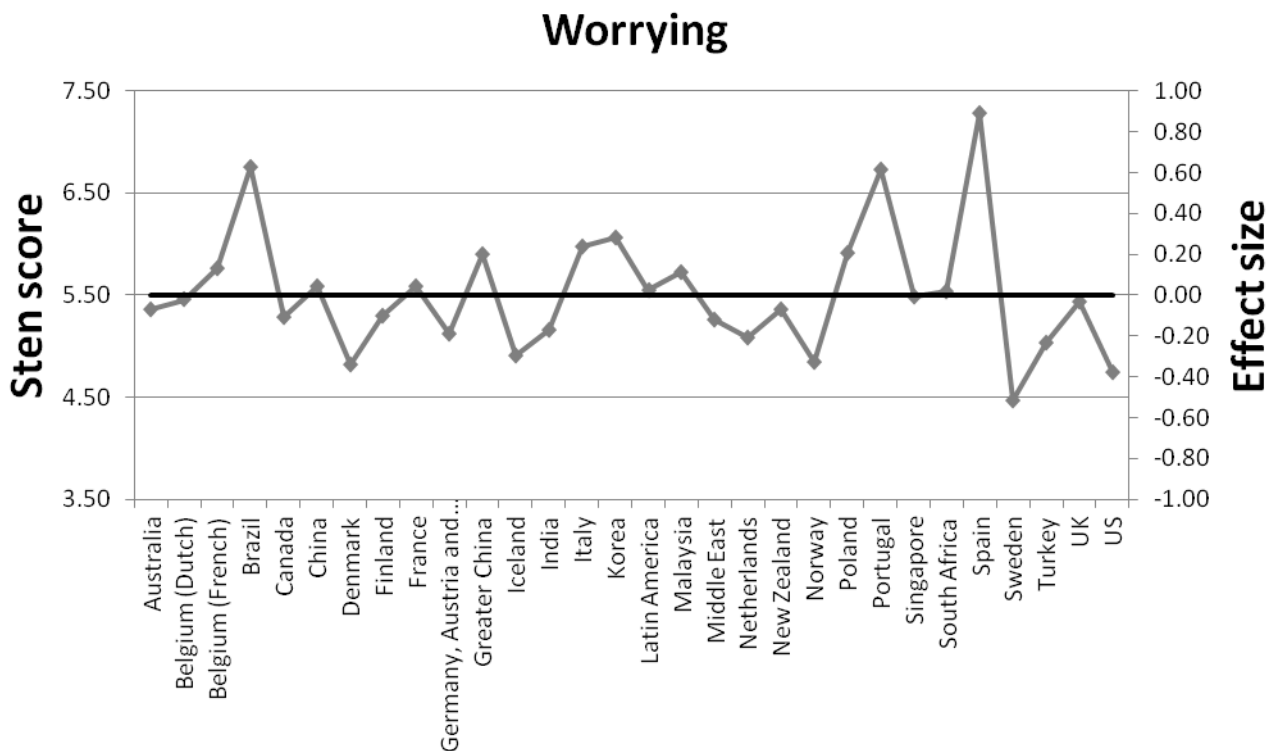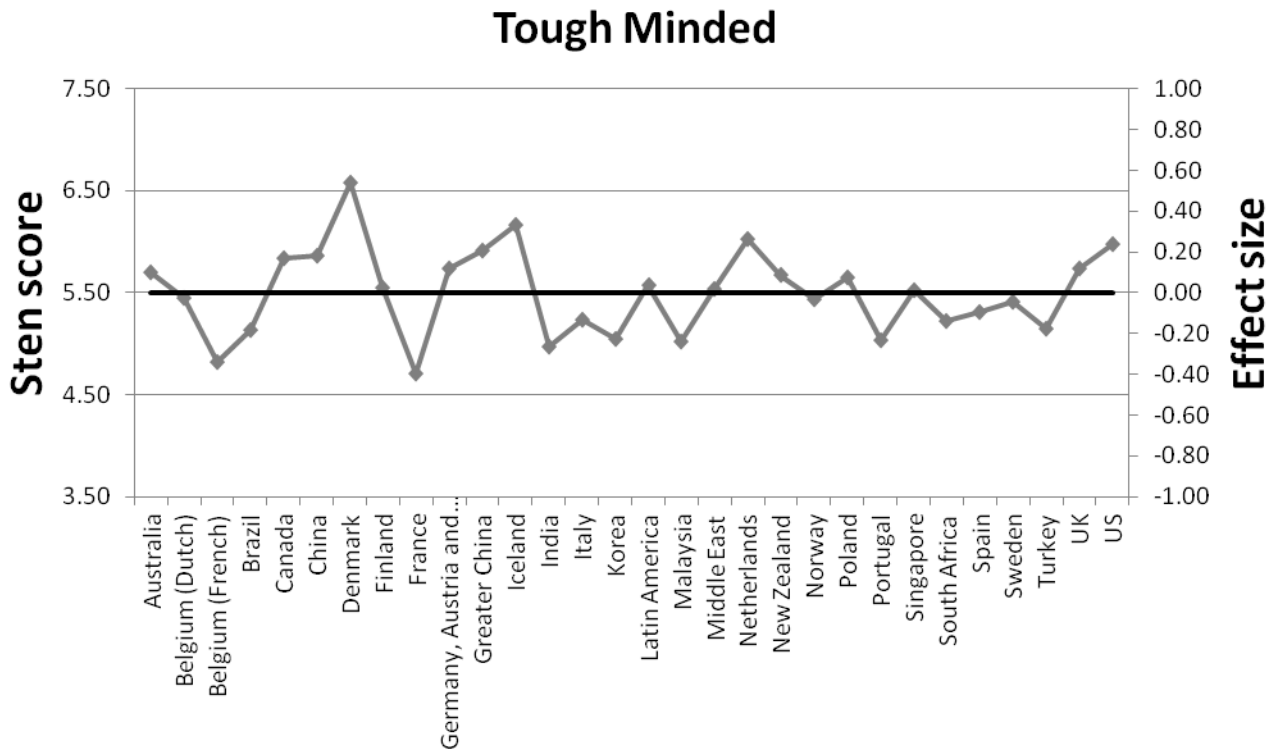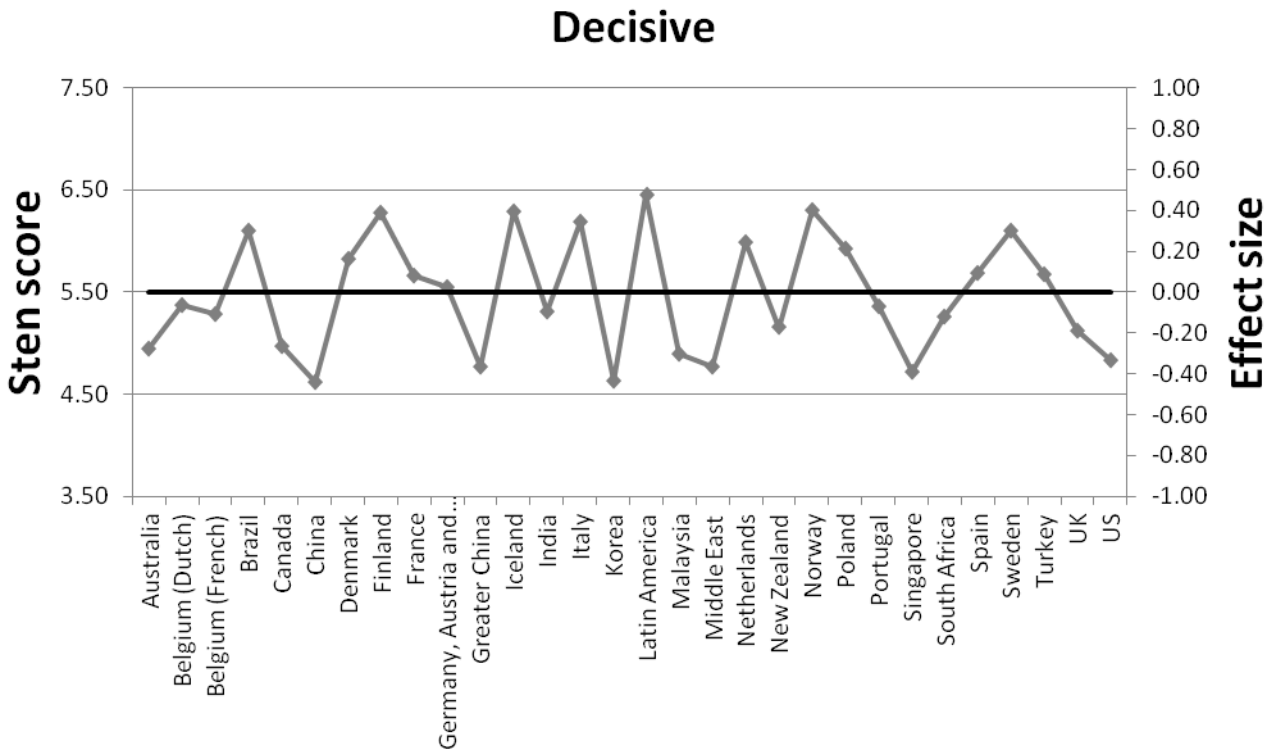
Chapter 8: Norms and comparison groups

Norms are part of the measurement procedure; they provide the *scaling* that is needed to assign a *value and meaning* to the raw scores obtained from an instrument. These raw scores are largely determined by a person's trait level. However, they also depend, to a greater or lesser degree (Bartram, 2008), on:

- Demographics:
    - o Endogenous factors: biological characteristics such as gender, age, or race
    - o Exogenous factors: environmental characteristics such as educational level and type, job level and type, organization, industrial sector, labour market, language, culture
- Temporal factors (for example, generational effects)
- Assessment factors including format (e.g. supervised session, remote administration) and setting (e.g. pre-screening, selection, development, research)

Any norm group (or reference or comparison group) can be thought of as reflecting a particular profile of these factors, and hence enables us to relate a person's score to a well-defined reference group. In practice, some of these factors may have little or no effect on scores. For example, mode of administration (i.e. online and paper & pencil) has been shown to have no significant effect on the psychometric properties of the OPQ32 (Bartram & Brown, 2004). In other cases factors such as demographic composition and culture might have more substantial or consistent effects on the expressed levels of a trait.

As described earlier, the output scores from OPQ32r are theta values. Theta refers to the scale score in IRT models that measure a latent trait, such as a personality construct, and typically ranges from -3 to +3. Theta based norms provide the means of converting these theta values into more familiar sten scores, where each scale has a mean of 5.5 and a standard deviation of 2.

When the OPQ32r was launched in 2009, all previously available OPQ32i norms were available through equating of the IRT-based and the classical test theory-based scores which are described in Section 8.2. All equated OPQ32i norms have since then been replaced with norms that were generated from OPQ32r theta data. Section 8.1 provides an overview of the UK and US OPQ32r general population norm groups. Detailed information on norms is available in the norm fact sheets and the technical documentation that was created for the norm updates (country and international norms: SHL 2015a,b,c).

8.1 OPQ32r norms

From 2011 to 2013 the OPQ32r norms were updated using OPQ32r theta data, resulting in over new 100 norms, spanning more than 24 languages and 40 countries/regions. These replaced the OPQ32i norms that were applied to the OPQ32r through equating procedures. All data was checked for normality, showing that OPQ32r theta scores were normally distributed. Norms were directly generated from OPQ32r theta score distributions using standard transformation procedures.

A norm should be representative of the candidate population and therefore consist of people with whom the candidate will be compared (Cronbach, 1990). The updated norms are based on data collected with individuals who completed the OPQ32r for assessment purposes in a selection or development context.

Table 45 summarises information on the distributions of the OPQ32r theta scores for the UK and US general population norm samples. Histograms showing the distribution of individual scales can be found in Appendix B.

.SHL.

Table 45: Distributions of the OPQ32r theta scores for the UK and US general population norm samples

	UK (N=22,612)				US (N=5,437)			
	Mean	**SD**	**Skewness**	**Kurtosis**	**Mean**	**SD**	**Skewness**	**Kurtosis**
Persuasive	0.79	0.87	-0.18	-0.19	1.19	0.88	-0.33	-0.11
Controlling	0.78	0.68	-0.33	0.29	0.97	0.63	-0.28	0.38
Outspoken	0.14	0.69	0.05	0.14	-0.05	0.66	0.16	-0.01
Independent Minded	-0.42	0.71	0.36	0.18	-0.39	0.70	0.39	0.30
Outgoing	0.16	0.73	0.10	0.07	0.34	0.68	0.03	0.19
Affiliative	0.05	0.68	0.05	0.22	0.11	0.66	0.09	0.13
Socially Confident	0.74	0.75	-0.57	0.32	0.95	0.68	-0.53	0.50
Modest	-0.23	0.72	0.09	0.00	-0.30	0.71	0.08	0.07
Democratic	0.56	0.81	-0.10	-0.17	0.18	0.80	0.09	-0.23
Caring	0.00	0.77	-0.08	0.03	-0.02	0.73	-0.05	0.08
Data Rational	0.66	0.88	-0.15	-0.40	0.67	0.84	-0.14	-0.31
Evaluative	0.73	0.75	-0.13	-0.21	0.48	0.73	0.09	-0.06
Behavioural	0.12	0.85	0.08	-0.26	0.14	0.81	0.07	-0.26
Conventional	-0.10	0.87	0.03	-0.23	0.00	0.81	0.02	-0.13
Conceptual	0.30	0.87	0.14	-0.10	0.23	0.85	0.22	-0.08
Innovative	0.71	0.82	-0.26	-0.06	0.78	0.76	-0.26	0.05
Variety Seeking	-0.11	0.86	0.36	-0.31	-0.21	0.81	0.40	-0.09
Adaptable	0.10	0.81	-0.05	-0.35	-0.16	0.78	0.21	-0.28
Forward Thinking	0.34	0.76	-0.09	0.12	0.51	0.69	-0.12	0.11
Detail Conscious	0.53	0.73	-0.24	0.25	0.46	0.72	-0.06	0.12
Conscientious	0.50	0.71	-0.38	0.39	0.50	0.65	-0.19	0.15
Rule Following	0.03	0.80	0.02	-0.13	0.30	0.81	-0.03	-0.12
Relaxed	0.24	0.71	-0.01	0.05	0.07	0.70	0.13	-0.03
Worrying	-0.74	0.80	0.45	-0.21	-1.04	0.70	0.61	0.19
Tough Minded	0.35	0.75	0.02	0.04	0.42	0.68	0.04	0.00
Optimistic	0.38	0.80	-0.20	0.16	0.65	0.78	-0.24	0.02
Trusting	0.49	0.65	-0.17	0.19	0.51	0.66	-0.18	0.16
Emotionally Controlled	-0.39	0.63	0.23	0.27	-0.50	0.57	0.14	0.19
Vigorous	0.85	0.61	-0.19	0.18	0.73	0.62	-0.14	0.06
Competitive	0.14	0.90	0.12	-0.42	0.80	0.95	-0.25	-0.42
Achieving	0.74	0.81	-0.26	-0.10	1.02	0.73	-0.42	0.26
Decisive	0.52	0.82	0.15	-0.17	0.39	0.77	0.36	0.13

Skewness indicates the extent to which a distribution deviates from symmetry around the mean. A value of zero means the distribution is symmetric, a positive value indicates a greater number of smaller values, and a negative value indicates a greater number of larger values. A skewness value of +/- 1 is considered very good for most psychometric uses, but +/- 2 is also usually acceptable. As can be seen in the table, the OPQ32r theta scores have skewness within +/- 1, suggesting that the distributions of the scores are very close to being symmetrical.

Kurtosis examines how peaked a distribution is compared to the normal distribution. A value of zero means the distribution's peak is closed to that of a normal distribution, a positive value indicates a distribution with a higher peak, and a negative value indicates a distribution with a flatter and wider peak. Like skewness, a kurtosis value of +/- 1 is considered very good for most psychometric uses, but +/- 2 is also usually acceptable. The OPQ32r theta scores have kurtosis within +/- 1 which, together with reasonably good skewness, suggests that the distributions of the scores are close to being normal.

The OPQ32r norms were created by following clearly defined standards such as appropriateness for use (whether local or international norms), appropriateness for intended applications, sample size and providing information on effects of age and gender and minority/protected group differences where appropriate, based on the European Federation of Psychologists Associations (EFPA) Test Review Criteria, Version 3.42 (EFPA, 2008). Sten scores were obtained directly from theta scores by normalising data using percentile cut points that divide the distribution into percentile bands corresponding to the appropriate standard score intervals. Details of the process and an overview of the norms are provided in the technical norm documentation (SHL, 2015c). In addition three international norms (general work population, managerial & professional and graduated) were created which are also documented (SHL, 2015b). All technical documentation and fact sheets that provide information on the language, country and demographic composition (e.g. gender, age, managerial level, ethnic background if appropriate) are available for each of the norm groups by contacting an account manager.

8.2 Equating the OPQ32r and OPQ32i

As the OPQ32r was created by removing one item out of each block of four, it is a shortened version of OPQ32i as far as the item content is concerned. It is also the shortened version of OPQ32i in relation to preference decisions made in each block, (i.e. the 312 pair-wise comparisons made when completing OPQ32r are a subset of the 624 pair-wise comparisons that would be performed to complete OPQ32i). It is easy to see when considering the six comparisons {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D} that a respondent has to make to rank-order four items A, B, C and D. When one item is removed from the block, the respondent no longer needs to perform three comparisons involving that item; however, the remaining comparisons are still performed, with the same outcomes. Therefore, responses to the full OPQ32i can be used to extract the reduced set of responses relevant to OPQ32r. When this reduced set of responses is scored using the IRT methodology with the established OPQ32r model parameters, the resulting score is equivalent to the relevant OPQ32r score.

This simple logic was used to equate OPQ32r scores to classical OPQ32i scores. The two sets of scores were based on responses given to one instrument (OPQ32i) in one administration, but using different methods of scoring as follows.

Using existing OPQ32i data, preference decisions were recoded into binary outcomes of pair-wise comparisons, and binary outcomes related to the no-longer-existing items removed. The remaining responses were then scored using the IRT methodology to produce the OPQ32r score. The OPQ32i classical (ipsative) score was computed as usual, from the full set of responses by summing the item scores as appropriate. A large online high-stakes sample was used to produce both OPQ32r and OPQ32i scores for equating.

Sample 8. OPQ32i high-stakes online sample. This is a large data extract from the SHL Talent Measurement online assessment systems. The OPQ32i in English was completed in real assessments in UK-based organisations by N=18,423 people. Of the sample, 61.6% were male, 37.3% female and 1.1% did not report their gender. Most participants described themselves as white (85.2%); the next largest groups were Asian (5.5%) and Black (2.2%). More information regarding the distribution of age can be found in Table 46.

Table 46: Distribution of age groups in sample 8.

Age band	Percentage of respondents
20 or under	0.6%
21 to 25	8.8%
26 to 30	14.3%
31 to 35	15.8%
36 to 40	17.0%
41 to 45	16.8%
46 to 50	13.4%
51+	11.0%
Not stated	2.3%

Table 47 below shows the correlations between classical test theory based scores for the full OPQ32i and the IRT-based OPQ32r scores. It can be seen that the OPQ32i provides very similar ordering of people on every scale. The correlations between scales range from 0.81 to 0.95 with a median of 0.90. The relationships between IRT-scored and classically scored forced-choice responses do not reach near-perfect correlations (0.97 and above; see Fan, 1998) that are typically found when comparing scores

derived from single-stimulus ratings by summing the item responses (CTT) and scoring with IRT. This is because the classical model does not adequately describe the process of responding to the forced-choice items, as we have seen, and this is where the IRT delivers real advantages. However, these relationships are sufficiently strong to be used as a basis for equating the raw ipsative OPQ32i score and the OPQ32r theta score. Linear equating was used to produce transformations necessary to convert the OPQ32r scores to the scale of raw OPQ32i scores. After the transformation was completed by the online scoring system, all legacy OPQ32i norms could be used when the OPQ32r was launched in 2009 before new OPQ32r theta based norms were available.

Table 47: Correlations between OPQ32i and OPQ32r scores derived from responses to OPQ32i.

	Correlations
Persuasive	0.91
Controlling	0.91
Outspoken	0.90
Independent Minded	0.81
Outgoing	0.91
Affiliative	0.86
Socially Confident	0.90
Modest	0.92
Democratic	0.89
Caring	0.86
Data Rational	0.94
Evaluative	0.84
Behavioural	0.90
Conventional	0.90
Conceptual	0.89
Innovative	0.93
Variety Seeking	0.83
Adaptable	0.94
Forward Thinking	0.88
Detail Conscious	0.88
Conscientious	0.88
Rule Following	0.92
Relaxed	0.90
Worrying	0.94
Tough Minded	0.88
Optimistic	0.86
Trusting	0.90
Emotionally Controlled	0.90
Vigorous	0.85
Competitive	0.95
Achieving	0.86
Decisive	0.90
Median	0.90

In summary, a wide range of theta-based norms are available for the OPQ32r. These theta based norms replaced the previous norms that were obtained through equating procedures to OPQ32i equivalent raw scores and which were then normed using existing OPQ32i norm tables.

References

Ackerman, T.A. (2005). Multidimensional Item Response Theory Modelling. In A. Maydeu-Olivares & J. J. McArdle. (Eds.), *Contemporary Psychometrics* (pp. 3-26). Mahwah, NJ: Lawrence Erlbaum.

Baron, H. (1996). Strengths and Limitations of Ipsative Measurement. *Journal of Occupational and Organisational Psychology, 69,* 49-56.

Bartram, D. (2013). Scalar equivalence of OPQ32: Big Five profiles of 31 countries. *Journal of Cross-Cultural Psychology 44(1*), 61-83.

Bartram, D. (2012). Stability of OPQ32 personality constructs across languages, culture and countries. In A. M. Ryan, F. T. L. Leong & F. L. Oswald (Eds). *Conducting Multinational Research: Applying Organizational Psychology in the Workplace*. Chapter 3, pp 59-90. Washington, D.C.: American Psychological Association .

Bartram, D. (2008). Global Norms? Towards some guidelines for aggregating personality norms across countries*. International Journal of Testing, 8*, 315-333.

Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment, 15*, 263-272.

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90,* 1185-1203.

Bartram, D. (1996). The relationship between ipsatised and normative measures of personality*. Journal of Occupational Psychology, 69*, 25-39.

Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ32i scores*. International Journal Of Selection And Assessment, 12(3)*: 278-284.

Bartram, D., Warr, P. & Brown, A. (2010). Let's focus on two-stage alignment, not just overall performance. *Industrial and Organisational Psychology, 3*, 335–339.

Bradley, K. & Hauenstein, N. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science, 48*, 313-335.

Brown, A. (2008). The Impact of Questionnaire Item Format on Ability to "Fake Good". In Brown, A. (chair): *Exploring the use of ipsative measures in personnel selection*. Symposium presented at the 6th Conference of the International Test Commission, Liverpool.

Brown, A. & Bartram, D. (2008). *IRT model for recovering latent traits from forced-choice personality tests*. Paper presented at the 23rd annual conference of the Society for Industrial and Organisational Psychology, San Francisco, CA.

Brown, A. & Maydeu-Olivares, A. (2011). Item response modelling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*, 460–502.

Brown, A., & Maydeu-Olivares, A. (2012). How IRT can solve problems of ipsative data in forced-choice questionnaires.In Byrne, B. M. (2006). *Structural Equation Modeling with EQS. Basic concepts, applications and programming* (2nd edition). New Jersey: Lawrence Erlbaum Associates, Publishers.

Byrne, B. M. (2006). *Structural Equation Modeling with EQS. Basic concepts, applications and programming* (2nd edition). New Jersey: Lawrence Erlbaum Associates, Publishers.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research, 1(2)*, 245-276.

Cheung, M.W.L, & Chan, W. (2002). Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 55-77.

Christiansen, N, Burns, G., & Montgomery, G. (2005). Reconsidering the use of forced-choice formats for applicant personality assessment. *Human Performance, 18*, 267-307.

Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14.*

Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational Psychology, 69*, 41-47.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th ed.). New York: Harper Row.

Dunlap, W. P., & Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research, 29*, 115-126.

EFPA (European Federation of Psychologists Associations, 2008). Test Review Criteria, Version 3.42.

Embretson, S. & Reise, S. (2000). *Item Response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fan, X. (1998). Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58(3)*, 357-381.

Friedman, H., & Amoo, T. (1999). Rating the rating scales. *Journal of Marketing Management, 9*, 114-123.

Funder, D.C. & Dobroth, K.M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52(2)*, 409-418.

Hogan, R., & Hogan, J. (1997). *Hogan Development Survey Manual*. Tulsa, OK: 401 Hogan Assessment Systems.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6(1)*, 1–55.

Jackson, D., Wroblewski, V., & Ashton, M. (2000). The Impact of Faking on Employment Tests: Does Forced Choice Offer a Solution*? Human Performance, 13*, 371–388.

Karpatschof, B., & Elkjaer, H. K. (2000). *Yet the bumblebee flies: The reliability of ipsative scores examined by empirical data and a simulation study*. Department of Psychology, University of Copenhagen: Research Report no. 1.

McCrae, R.R.; Costa, P.T.; Jr (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52* (1), 81–90

Martin, B. A., Bowen C.C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247-256.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika, 64*, 325-340.

Maydeu-Olivares, A. & Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods, 10*, 285-304.

Maydeu-Olivares, A. & Brown, A. (2010). Item response modeling of paired comparison and ranking data. *Multivariate Behavioural Research, 45*, 935-974.

McCloy, R., Heggestad, E., Reeve, C. (2005). A Silk Purse From the Sow,s Ear: Retrieving Normative Information From Multidimensional Forced-Choice Items. *Organisational Research Methods, 8,* 222-248.

Meade, A. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organisational Psychology, 77*, 531-552.

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology, 78*, 218-225.

Muthén, L.K. & Muthén, B.O. (1998-2007). *Mplus User's guide*. Fifth edition. Los Angeles, CA: Muthén & Muthén.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer

SHL. (2006)*. OPQ32 Technical Manual*. Thames Ditton, UK: SHL.

SHL. (2014). *Motivation Questionnaire manual and user's guide*. Surrey, UK: SHL.

SHL. (2015a). *OPQ32r Consistency Score User Guidance*. Surrey, UK. SHL.

SHL. (2015b). *OPQ International Norm Technical Manual*. Surrey, UK. SHL.

SHL. (2015c). *OPQ32r Norm Update Technical Documentation*. Surrey, UK. SHL.

Stark, S., Chernyshenko, O. & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The Multi-Unidimensional Pairwise-Preference Model. *Applied Psychological Measurement, 29*, 184-203.

Tenopyr, M. L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology, 73*, 749-751.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review, 79*, 281-299.

Thurstone, L.L. (1931). Rank order as a psychological method. *Journal of Experimental Psychology, 14*, 187-201.

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* (pp. 39-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Van Herk, H., Poortinga, Y., & Verhallen, T. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35*, 346.

Warr, P. B. (1999). Logical and judgmental moderators of the criterion-related validity of personality scales. *Journal of Occupational and Organizational Psychology, 72*, 187-204.

Appendix A – Sample description and IRT composite reliabilities and standard errors

Table 48: Distribution of gender, age and education in the international norm sample.

Gender	N	%
Male	72,634	61.4
Female	45,690	38.6
Total	118,324	100.0
Age	**N**	**%**
Under 18	24	.0
18-20	1,383	1.7
21-24	11,748	14.9
25-29	13,799	17.5
30-34	13,178	16.7
35-39	13,350	16.9
40-44	11,240	14.2
45-49	7,646	9.7
50-54	4,520	5.7
55-59	1,790	2.3
60-64	353	.4
65 or older	32	.0
Total	79,063	100.0
Education	**N**	**%**
No Qualification	541	1.0
Secondary/High school	11,824	22.7
Vocational	6,459	12.4
Bachelor	17,957	34.5
Master	14,563	28.0
Doctorate	755	1.4
Total	52,099	100.0

Job level, Managerial role and Industry

Table 49: Distribution of job level, managerial role and industry in the international norm sample.

Job level	N	%
Executive	2,511	5.6
Senior Manager	3,508	7.9
Manager	7,737	17.3
Supervisor/Team leader	3,761	8.4
Employee	17,969	40.3
Trainee	606	1.4
Self employed	1,462	3.3
Not employed	7,071	15.8
Total	44,625	100.0
Managerial	**N**	**%**
Non-managerial	27,108	60.7
Managerial (incl. team leader)	17,517	39.3
Total	44,625	100.0
Industry	**N**	**%**
Consulting and Professional Services	43,966	37.2
Finance and Insurance	18,095	15.3
Technology and Telecoms	11,366	9.6
Education, Government, Healthcare, Non Profit	11,786	10.0
Consumer Services	12,739	10.8
Manufacturing, Construction, Transportation and Utilities	20,372	17.2
Total	118,324	100.0

Table 50: Overview of countries included and samples sizes in the international norm sample.

Regions	Countries included	N	%
Australia	Australia	9,120	7.7
Belgium (Dutch)	Belgium	2,385	2.0
Belgium (French)	Belgium	2,529	2.1
Brazil	Brazil	1,006	.9
Canada	Canada	703	.6
China	China	3,322	2.8
Denmark	Denmark	6,809	5.8
Finland	Finland	5,381	4.5
France	France	4,225	3.6
Germany, Austria and Switzerland	Austria, Germany, Switzerland	2,336	2.0
Greater China	Hong Kong, Taiwan	564	.5
Iceland	Iceland	729	.6
India	India	1,098	.9
Italy	Italy	4,550	3.8
Korea	Korea	846	.7
Malaysia	Malaysia	2,672	2.3
Middle East	Bahrain, Egypt, Kuwait, Saudi Arabia, United Arab Emirates	1,494	1.3
Netherlands	Netherlands	6,227	5.3
New Zealand	New Zealand	2,818	2.4
Norway	Norway	4,659	3.9
Latin-America	Argentina, Chile, Colombia, Costa Rica, El Salvador, Honduras, Mexico, Venezuela	971	.8
Poland	Poland	927	.8
Portugal	Portugal	2,248	1.9
Singapore	Singapore	4,053	3.4
South Africa	South Africa	4,880	4.1
Spain	Spain	695	.6
Sweden	Sweden	13,244	11.2
Turkey	Turkey	1,107	.9
UK	United Kingdom	22,612	19.1
US	United States	4,114	3.5
Total		118,324	100.0

Table 51: Reliabilities and standard errors by country in the international norm sample.

	Australia		Belgium (Dutch)		Belgium (French)		Brazil		Canada	
N	**9,120**		**2,385**		**2,529**		**1,006**		**703**	
	Rel.	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**
Persuasive	0.84	0.39	0.84	0.40	0.84	0.39	0.86	0.37	0.85	0.38
Controlling	0.93	0.26	0.92	0.27	0.93	0.26	0.93	0.26	0.93	0.26
Outspoken	0.87	0.36	0.86	0.37	0.87	0.36	0.87	0.35	0.87	0.36
Independent minded	0.76	0.49	0.76	0.49	0.76	0.49	0.77	0.48	0.76	0.49
Outgoing	0.88	0.33	0.89	0.33	0.90	0.32	0.90	0.31	0.89	0.33
Affiliative	0.86	0.37	0.85	0.39	0.86	0.37	0.88	0.35	0.87	0.36
Socially confident	0.88	0.34	0.88	0.34	0.89	0.32	0.90	0.32	0.88	0.34
Modest	0.82	0.41	0.82	0.42	0.83	0.40	0.83	0.41	0.82	0.42
Democratic	0.75	0.50	0.75	0.50	0.76	0.49	0.77	0.48	0.75	0.50
Caring	0.83	0.41	0.83	0.41	0.83	0.41	0.85	0.39	0.83	0.41
Data rational	0.89	0.32	0.89	0.32	0.91	0.30	0.91	0.30	0.89	0.32
Evaluative	0.79	0.45	0.78	0.46	0.79	0.46	0.79	0.45	0.79	0.45
Behavioural	0.80	0.44	0.80	0.44	0.81	0.44	0.81	0.43	0.80	0.44
Conventional	0.68	0.56	0.68	0.56	0.69	0.55	0.70	0.54	0.68	0.56
Conceptual	0.78	0.47	0.78	0.47	0.79	0.46	0.79	0.46	0.78	0.47
Innovative	0.89	0.33	0.89	0.33	0.89	0.32	0.90	0.31	0.89	0.33
Variety seeking	0.77	0.48	0.77	0.48	0.78	0.47	0.79	0.46	0.76	0.48
Adaptable	0.87	0.35	0.86	0.36	0.89	0.33	0.88	0.34	0.87	0.35
Forward thinking	0.88	0.34	0.88	0.35	0.88	0.34	0.89	0.33	0.87	0.35
Detail conscious	0.89	0.33	0.89	0.32	0.89	0.33	0.90	0.31	0.89	0.32
Conscientious	0.83	0.41	0.82	0.42	0.82	0.42	0.86	0.37	0.83	0.40
Rule following	0.89	0.32	0.89	0.32	0.89	0.32	0.90	0.31	0.89	0.32
Relaxed	0.89	0.33	0.88	0.34	0.89	0.33	0.90	0.32	0.88	0.34
Worrying	0.77	0.47	0.77	0.47	0.79	0.45	0.82	0.42	0.76	0.48
Tough minded	0.80	0.44	0.80	0.45	0.81	0.44	0.81	0.43	0.81	0.44
Optimistic	0.81	0.43	0.81	0.43	0.82	0.43	0.82	0.42	0.81	0.43
Trusting	0.90	0.31	0.90	0.32	0.90	0.31	0.90	0.31	0.90	0.31
Emotionally controlled	0.87	0.35	0.86	0.37	0.88	0.34	0.87	0.35	0.87	0.36
Vigorous	0.89	0.33	0.89	0.32	0.90	0.31	0.90	0.31	0.89	0.33
Competitive	0.86	0.37	0.85	0.38	0.87	0.35	0.89	0.34	0.87	0.36
Achieving	0.78	0.46	0.79	0.46	0.80	0.45	0.79	0.45	0.78	0.47
Decisive	0.84	0.39	0.85	0.38	0.86	0.38	0.86	0.37	0.84	0.39
Median	0.85	0.38	0.85	0.38	0.86	0.37	0.87	0.36	0.86	0.37
Average	0.84	0.39	0.84	0.40	0.84	0.38	0.85	0.38	0.84	0.39
Min	0.68	0.26	0.68	0.27	0.69	0.26	0.70	0.26	0.68	0.26
Max	0.93	0.56	0.92	0.56	0.93	0.55	0.93	0.54	0.93	0.56

	China		Denmark		Finland		France		Germany, Austria and Switzerland	
N	**3,322**		**6,809**		**5,381**		**4,225**		**2,336**	
	Rel.	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**	**Rel.**	**SE.**
Persuasive	0.85	0.38	0.85	0.39	0.85	0.39	0.85	0.38	0.86	0.38
Controlling	0.93	0.25	0.92	0.27	0.92	0.27	0.93	0.26	0.93	0.25
Outspoken	0.87	0.36	0.87	0.36	0.87	0.36	0.87	0.35	0.87	0.36
Independent minded	0.79	0.45	0.77	0.48	0.77	0.48	0.76	0.49	0.78	0.47
Outgoing	0.89	0.33	0.90	0.31	0.89	0.32	0.90	0.31	0.90	0.32
Affiliative	0.87	0.36	0.87	0.36	0.86	0.37	0.87	0.36	0.87	0.36
Socially confident	0.89	0.33	0.89	0.33	0.88	0.34	0.90	0.32	0.89	0.33
Modest	0.84	0.40	0.80	0.43	0.82	0.42	0.84	0.39	0.82	0.41
Democratic	0.75	0.50	0.75	0.50	0.75	0.49	0.76	0.48	0.77	0.48
Caring	0.84	0.40	0.84	0.40	0.83	0.41	0.84	0.40	0.84	0.40
Data rational	0.91	0.30	0.90	0.31	0.90	0.32	0.91	0.30	0.91	0.30
Evaluative	0.81	0.43	0.80	0.45	0.80	0.45	0.79	0.45	0.81	0.44
Behavioural	0.82	0.43	0.81	0.44	0.81	0.44	0.81	0.44	0.81	0.43
Conventional	0.69	0.56	0.68	0.56	0.68	0.56	0.69	0.55	0.68	0.56
Conceptual	0.80	0.45	0.78	0.47	0.78	0.47	0.79	0.46	0.79	0.45
Innovative	0.89	0.32	0.89	0.32	0.89	0.33	0.90	0.32	0.89	0.33
Variety seeking	0.78	0.47	0.78	0.47	0.78	0.47	0.78	0.46	0.79	0.46
Adaptable	0.89	0.33	0.87	0.35	0.87	0.36	0.89	0.33	0.89	0.33
Forward thinking	0.86	0.37	0.88	0.34	0.88	0.35	0.89	0.33	0.89	0.33
Detail conscious	0.90	0.30	0.89	0.32	0.90	0.31	0.89	0.32	0.91	0.30
Conscientious	0.84	0.39	0.83	0.41	0.84	0.40	0.83	0.40	0.84	0.39
Rule following	0.89	0.32	0.90	0.32	0.90	0.32	0.90	0.31	0.91	0.31
Relaxed	0.89	0.32	0.90	0.32	0.89	0.33	0.89	0.33	0.90	0.32
Worrying	0.79	0.45	0.74	0.50	0.77	0.47	0.79	0.45	0.76	0.48
Tough minded	0.82	0.42	0.82	0.42	0.81	0.43	0.80	0.44	0.81	0.43
Optimistic	0.81	0.43	0.81	0.43	0.80	0.45	0.82	0.42	0.81	0.43
Trusting	0.90	0.31	0.90	0.31	0.90	0.31	0.91	0.31	0.90	0.31
Emotionally controlled	0.88	0.34	0.87	0.36	0.86	0.36	0.89	0.33	0.88	0.35
Vigorous	0.89	0.32	0.88	0.34	0.89	0.33	0.91	0.31	0.90	0.31
Competitive	0.89	0.33	0.86	0.36	0.86	0.37	0.87	0.36	0.87	0.36
Achieving	0.80	0.45	0.79	0.46	0.79	0.46	0.79	0.45	0.79	0.46
Decisive	0.83	0.40	0.86	0.37	0.86	0.37	0.86	0.37	0.87	0.36
Median	0.86	0.37	0.86	0.37	0.86	0.37	0.86	0.37	0.87	0.36
Average	0.85	0.38	0.84	0.39	0.84	0.39	0.85	0.38	0.85	0.38
Min	0.69	0.25	0.68	0.27	0.68	0.27	0.69	0.26	0.68	0.25
Max	0.93	0.56	0.92	0.56	0.92	0.56	0.93	0.55	0.93	0.56

N	Greater China		Iceland		India		Italy		Korea	
	564		729		1,098		4,550		846	
	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.
Persuasive	0.86	0.37	0.84	0.39	0.85	0.39	0.85	0.39	0.85	0.38
Controlling	0.93	0.26	0.93	0.26	0.93	0.26	0.93	0.26	0.94	0.24
Outspoken	0.87	0.36	0.87	0.35	0.87	0.36	0.87	0.36	0.88	0.35
Independent minded	0.79	0.46	0.78	0.47	0.77	0.47	0.78	0.47	0.76	0.49
Outgoing	0.90	0.31	0.88	0.34	0.88	0.33	0.90	0.31	0.90	0.31
Affiliative	0.86	0.37	0.87	0.36	0.87	0.36	0.87	0.36	0.84	0.39
Socially confident	0.89	0.32	0.89	0.32	0.89	0.33	0.89	0.33	0.89	0.33
Modest	0.83	0.40	0.83	0.40	0.82	0.41	0.83	0.41	0.83	0.40
Democratic	0.75	0.50	0.75	0.50	0.75	0.49	0.76	0.49	0.76	0.49
Caring	0.84	0.40	0.84	0.40	0.84	0.40	0.84	0.40	0.84	0.39
Data rational	0.90	0.31	0.90	0.31	0.90	0.31	0.91	0.30	0.90	0.31
Evaluative	0.81	0.44	0.80	0.45	0.81	0.44	0.79	0.45	0.81	0.44
Behavioural	0.81	0.43	0.80	0.44	0.81	0.43	0.80	0.44	0.82	0.42
Conventional	0.68	0.56	0.69	0.55	0.68	0.56	0.68	0.56	0.70	0.55
Conceptual	0.79	0.45	0.78	0.47	0.79	0.46	0.79	0.46	0.79	0.46
Innovative	0.89	0.32	0.89	0.33	0.89	0.33	0.90	0.32	0.90	0.32
Variety seeking	0.78	0.47	0.79	0.45	0.76	0.48	0.78	0.46	0.79	0.45
Adaptable	0.88	0.34	0.86	0.37	0.88	0.34	0.88	0.34	0.89	0.32
Forward thinking	0.87	0.36	0.88	0.34	0.87	0.36	0.89	0.33	0.88	0.34
Detail conscious	0.91	0.30	0.89	0.32	0.90	0.31	0.91	0.30	0.91	0.30
Conscientious	0.85	0.38	0.84	0.39	0.85	0.39	0.85	0.39	0.86	0.36
Rule following	0.90	0.31	0.90	0.32	0.89	0.32	0.90	0.32	0.90	0.32
Relaxed	0.89	0.32	0.89	0.33	0.89	0.33	0.89	0.33	0.90	0.32
Worrying	0.80	0.44	0.76	0.48	0.77	0.47	0.80	0.44	0.80	0.44
Tough minded	0.82	0.42	0.82	0.43	0.81	0.43	0.81	0.43	0.82	0.42
Optimistic	0.82	0.43	0.81	0.43	0.80	0.44	0.82	0.42	0.82	0.43
Trusting	0.90	0.32	0.90	0.31	0.89	0.32	0.89	0.32	0.90	0.31
Emotionally controlled	0.88	0.35	0.87	0.36	0.88	0.34	0.88	0.34	0.88	0.35
Vigorous	0.89	0.32	0.87	0.35	0.89	0.32	0.90	0.32	0.89	0.32
Competitive	0.88	0.34	0.85	0.38	0.89	0.33	0.88	0.34	0.88	0.34
Achieving	0.80	0.45	0.79	0.46	0.78	0.47	0.79	0.45	0.80	0.44
Decisive	0.84	0.39	0.86	0.38	0.86	0.37	0.86	0.37	0.84	0.40
Median	0.86	0.37	0.85	0.38	0.86	0.37	0.86	0.37	0.86	0.37
Average	0.85	0.38	0.84	0.39	0.84	0.39	0.85	0.38	0.85	0.38
Min	0.68	0.26	0.69	0.26	0.68	0.26	0.68	0.26	0.70	0.24
Max	0.93	0.56	0.93	0.55	0.93	0.56	0.93	0.56	0.94	0.55

	Malaysia		Middle East		Netherlands		New Zealand		Norway	
N	**2,672**		**1,494**		**6,227**		**2,818**		**4,659**	
	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.
Persuasive	0.85	0.38	0.86	0.37	0.84	0.39	0.84	0.39	0.84	0.39
Controlling	0.94	0.25	0.94	0.25	0.93	0.27	0.93	0.26	0.92	0.27
Outspoken	0.88	0.35	0.88	0.35	0.87	0.36	0.87	0.36	0.87	0.36
Independent minded	0.78	0.47	0.78	0.46	0.77	0.48	0.76	0.49	0.74	0.51
Outgoing	0.90	0.31	0.90	0.31	0.90	0.32	0.88	0.34	0.89	0.32
Affiliative	0.87	0.35	0.87	0.36	0.85	0.38	0.86	0.37	0.86	0.37
Socially confident	0.90	0.32	0.90	0.32	0.89	0.33	0.88	0.33	0.89	0.33
Modest	0.84	0.40	0.85	0.39	0.82	0.42	0.83	0.41	0.82	0.41
Democratic	0.77	0.48	0.76	0.48	0.76	0.49	0.75	0.50	0.76	0.49
Caring	0.84	0.40	0.85	0.39	0.83	0.41	0.83	0.41	0.84	0.40
Data rational	0.90	0.30	0.91	0.30	0.89	0.32	0.89	0.32	0.91	0.30
Evaluative	0.81	0.43	0.82	0.42	0.80	0.45	0.79	0.45	0.80	0.45
Behavioural	0.82	0.43	0.82	0.42	0.80	0.44	0.81	0.44	0.81	0.43
Conventional	0.70	0.55	0.71	0.54	0.68	0.56	0.68	0.56	0.69	0.56
Conceptual	0.80	0.45	0.80	0.45	0.78	0.47	0.78	0.47	0.78	0.47
Innovative	0.90	0.31	0.90	0.31	0.89	0.33	0.89	0.33	0.89	0.32
Variety seeking	0.78	0.46	0.79	0.46	0.78	0.47	0.77	0.48	0.79	0.46
Adaptable	0.89	0.33	0.90	0.32	0.87	0.35	0.87	0.36	0.87	0.36
Forward thinking	0.87	0.35	0.88	0.34	0.88	0.34	0.88	0.34	0.88	0.34
Detail conscious	0.90	0.31	0.90	0.31	0.90	0.32	0.88	0.33	0.90	0.31
Conscientious	0.86	0.37	0.84	0.39	0.85	0.38	0.82	0.41	0.83	0.41
Rule following	0.90	0.31	0.88	0.34	0.89	0.33	0.89	0.32	0.90	0.31
Relaxed	0.89	0.33	0.89	0.32	0.89	0.33	0.89	0.33	0.90	0.32
Worrying	0.81	0.44	0.80	0.45	0.75	0.48	0.77	0.47	0.75	0.49
Tough minded	0.82	0.43	0.82	0.42	0.81	0.43	0.80	0.44	0.81	0.44
Optimistic	0.82	0.42	0.82	0.42	0.81	0.43	0.81	0.43	0.81	0.43
Trusting	0.90	0.32	0.90	0.31	0.90	0.31	0.90	0.31	0.90	0.31
Emotionally controlled	0.88	0.34	0.89	0.33	0.86	0.37	0.87	0.35	0.87	0.36
Vigorous	0.90	0.32	0.90	0.31	0.89	0.32	0.88	0.33	0.90	0.32
Competitive	0.89	0.34	0.89	0.33	0.86	0.37	0.86	0.37	0.87	0.36
Achieving	0.79	0.45	0.79	0.46	0.79	0.46	0.79	0.46	0.78	0.46
Decisive	0.86	0.37	0.85	0.38	0.86	0.37	0.85	0.39	0.86	0.37
Median	0.86	0.36	0.86	0.37	0.85	0.38	0.85	0.38	0.86	0.37
Average	0.85	0.38	0.85	0.37	0.84	0.39	0.84	0.39	0.84	0.39
Min	0.70	0.25	0.71	0.25	0.68	0.27	0.68	0.26	0.69	0.27
Max	0.94	0.55	0.94	0.54	0.93	0.56	0.93	0.56	0.92	0.56

	Latin America		Poland		Portugal		Singapore		South Africa	
N	**971**		**927**		**2,248**		**4,053**		**4,880**	
	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.
Persuasive	0.86	0.38	0.86	0.38	0.84	0.40	0.84	0.40	0.85	0.38
Controlling	0.92	0.28	0.93	0.26	0.93	0.26	0.93	0.26	0.93	0.26
Outspoken	0.86	0.37	0.87	0.36	0.87	0.36	0.87	0.36	0.87	0.36
Independent minded	0.76	0.48	0.78	0.47	0.76	0.49	0.76	0.48	0.78	0.47
Outgoing	0.89	0.32	0.89	0.33	0.89	0.33	0.89	0.33	0.89	0.33
Affiliative	0.88	0.35	0.87	0.36	0.87	0.36	0.86	0.37	0.87	0.36
Socially confident	0.90	0.31	0.88	0.34	0.89	0.33	0.88	0.33	0.90	0.32
Modest	0.82	0.42	0.83	0.41	0.83	0.41	0.82	0.41	0.83	0.40
Democratic	0.75	0.50	0.76	0.49	0.76	0.49	0.75	0.50	0.76	0.49
Caring	0.84	0.40	0.84	0.40	0.84	0.40	0.83	0.41	0.84	0.40
Data rational	0.90	0.31	0.90	0.32	0.90	0.31	0.90	0.32	0.90	0.31
Evaluative	0.80	0.45	0.79	0.45	0.78	0.46	0.80	0.45	0.80	0.44
Behavioural	0.82	0.42	0.81	0.44	0.81	0.44	0.80	0.44	0.82	0.43
Conventional	0.69	0.55	0.69	0.56	0.70	0.55	0.68	0.56	0.69	0.55
Conceptual	0.79	0.45	0.78	0.46	0.79	0.46	0.78	0.46	0.79	0.45
Innovative	0.89	0.33	0.89	0.33	0.90	0.32	0.89	0.33	0.90	0.32
Variety seeking	0.77	0.48	0.79	0.46	0.78	0.46	0.77	0.47	0.78	0.47
Adaptable	0.84	0.38	0.88	0.34	0.87	0.35	0.88	0.35	0.87	0.35
Forward thinking	0.88	0.34	0.88	0.34	0.88	0.34	0.86	0.36	0.88	0.34
Detail conscious	0.90	0.31	0.89	0.32	0.88	0.33	0.89	0.32	0.89	0.32
Conscientious	0.84	0.40	0.85	0.38	0.83	0.41	0.83	0.41	0.83	0.40
Rule following	0.90	0.31	0.90	0.32	0.90	0.31	0.90	0.32	0.90	0.32
Relaxed	0.88	0.34	0.89	0.33	0.89	0.33	0.89	0.33	0.89	0.33
Worrying	0.78	0.46	0.79	0.45	0.81	0.43	0.78	0.46	0.79	0.45
Tough minded	0.82	0.42	0.81	0.43	0.81	0.44	0.81	0.43	0.81	0.43
Optimistic	0.82	0.42	0.82	0.42	0.82	0.42	0.81	0.43	0.82	0.43
Trusting	0.90	0.32	0.90	0.32	0.90	0.32	0.89	0.32	0.89	0.32
Emotionally controlled	0.87	0.36	0.88	0.35	0.87	0.36	0.87	0.35	0.88	0.34
Vigorous	0.89	0.32	0.89	0.33	0.90	0.31	0.89	0.32	0.90	0.32
Competitive	0.88	0.34	0.87	0.35	0.88	0.34	0.87	0.36	0.88	0.34
Achieving	0.78	0.47	0.79	0.45	0.78	0.46	0.79	0.46	0.79	0.46
Decisive	0.86	0.37	0.85	0.38	0.85	0.38	0.84	0.40	0.86	0.37
Median	0.86	0.37	0.86	0.37	0.86	0.37	0.85	0.38	0.86	0.37
Average	0.84	0.39	0.84	0.39	0.84	0.39	0.84	0.39	0.85	0.38
Min	0.69	0.28	0.69	0.26	0.70	0.26	0.68	0.26	0.69	0.26
Max	0.92	0.55	0.93	0.56	0.93	0.55	0.93	0.56	0.93	0.55

	Spain		Sweden		Turkey		UK		US	
N	695		13,244		1,107		22,612		4,114	
	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.	Rel.	SE.
Persuasive	0.85	0.38	0.85	0.38	0.85	0.38	0.85	0.38	0.85	0.38
Controlling	0.94	0.25	0.93	0.26	0.92	0.28	0.93	0.26	0.92	0.27
Outspoken	0.88	0.35	0.87	0.36	0.86	0.38	0.87	0.36	0.87	0.37
Independent minded	0.77	0.48	0.76	0.49	0.76	0.49	0.76	0.49	0.76	0.48
Outgoing	0.89	0.32	0.90	0.31	0.90	0.32	0.89	0.33	0.89	0.32
Affiliative	0.87	0.36	0.86	0.37	0.87	0.36	0.87	0.36	0.87	0.36
Socially confident	0.89	0.33	0.89	0.33	0.90	0.32	0.88	0.34	0.88	0.34
Modest	0.83	0.41	0.82	0.41	0.78	0.45	0.82	0.42	0.82	0.42
Democratic	0.76	0.49	0.75	0.50	0.76	0.48	0.75	0.50	0.75	0.50
Caring	0.84	0.39	0.84	0.40	0.83	0.41	0.83	0.41	0.83	0.41
Data rational	0.90	0.31	0.90	0.31	0.89	0.32	0.89	0.32	0.89	0.32
Evaluative	0.79	0.46	0.80	0.45	0.77	0.47	0.79	0.46	0.80	0.44
Behavioural	0.81	0.43	0.80	0.44	0.81	0.44	0.80	0.44	0.80	0.44
Conventional	0.70	0.55	0.68	0.56	0.68	0.56	0.68	0.56	0.69	0.56
Conceptual	0.79	0.46	0.78	0.47	0.78	0.46	0.78	0.47	0.78	0.47
Innovative	0.90	0.32	0.89	0.33	0.88	0.34	0.89	0.33	0.89	0.33
Variety seeking	0.79	0.46	0.78	0.47	0.77	0.47	0.77	0.48	0.76	0.48
Adaptable	0.88	0.34	0.88	0.34	0.85	0.38	0.87	0.35	0.86	0.37
Forward thinking	0.89	0.33	0.89	0.33	0.87	0.35	0.88	0.34	0.88	0.35
Detail conscious	0.90	0.32	0.89	0.32	0.90	0.31	0.88	0.33	0.89	0.32
Conscientious	0.82	0.41	0.83	0.41	0.82	0.42	0.82	0.41	0.83	0.41
Rule following	0.90	0.31	0.90	0.32	0.90	0.31	0.89	0.32	0.89	0.32
Relaxed	0.89	0.33	0.90	0.32	0.89	0.32	0.89	0.33	0.88	0.34
Worrying	0.82	0.42	0.73	0.51	0.75	0.49	0.77	0.47	0.76	0.48
Tough minded	0.81	0.43	0.81	0.44	0.80	0.44	0.80	0.44	0.81	0.43
Optimistic	0.83	0.42	0.81	0.43	0.82	0.42	0.82	0.43	0.81	0.44
Trusting	0.90	0.31	0.90	0.31	0.87	0.35	0.90	0.31	0.90	0.32
Emotionally controlled	0.88	0.35	0.86	0.37	0.86	0.37	0.87	0.35	0.87	0.35
Vigorous	0.90	0.32	0.88	0.34	0.89	0.32	0.89	0.33	0.89	0.33
Competitive	0.88	0.35	0.86	0.36	0.88	0.34	0.86	0.36	0.86	0.36
Achieving	0.79	0.45	0.79	0.46	0.78	0.47	0.78	0.47	0.77	0.48
Decisive	0.86	0.37	0.86	0.37	0.86	0.38	0.85	0.39	0.84	0.39
Median	0.86	0.37	0.86	0.37	0.85	0.38	0.86	0.37	0.86	0.37
Average	0.85	0.38	0.84	0.39	0.84	0.39	0.84	0.39	0.84	0.39
Min	0.70	0.25	0.68	0.26	0.68	0.28	0.68	0.26	0.69	0.27
Max	0.94	0.55	0.93	0.56	0.92	0.56	0.93	0.56	0.92	0.56

Appendix B – Distribution of theta scores

OPQ Persuasive theta score

OPQ Controlling theta score

OPQ Outspoken theta score

OPQ Independent Minded theta score

OPQ Outgoing theta score

OPQ Affiliative theta score