The Mini-IPIP6: Item Response Theory analysis of a short measure of the big-six factors of personality in New Zealand

Chris G. Sibley, *University of Auckland*

This paper is the second in a series documenting and validating the Mini-IPIP6 for use in New Zealand. The Mini-IPIP6 is a public domain short-form personality instrument which provides four-item markers of the six broad-bandwidth dimensions of personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness to Experience and Honesty-Humility. This study reports results from an Item Response Theory analysis of the Mini-IPIP6 in a nationally representative New Zealand sample (N = 5,576). A series of unidimensional graded item response models indicated that the Mini-IPIP6 provided well distributed estimates of each of the six dimensions of personality across the latent trait range and centered on the population mean. These findings indicate that the Mini-IPIP6 provides a reasonably precise measure of the major dimensions of personality, given the scale's brevity. Discrimination and difficulty parameter estimates for the Mini-IPIP6 in the New Zealand population are presented, along with Test Information Functions for each personality dimension. A copy of the Mini-IPIP6 is included.

There are a vast number of self-report personality measures available. Generally speaking, these measures are based on the assumption that "… individuals are characterized by stable, distinctive, and highly meaningful patterns of variability in their actions, thoughts, and feelings across different types of situations. These *if … then …* situation-behavior relationships provide a kind of 'behavioral signature of personality' that identifies the individual and maps on to the impressions formed by observers about what they are like" (Mischel, 2004, p. 8). This quote provides a good working definition of personality. The aim of the many available personality measures should be then to provide a method for measuring individual differences in these distinct and highly meaningful patterns of variation, differences in other words, across people in their personality traits.

This study is the second in a series of manuscripts validating a short-form six-factor self-report measure of the six major dimensions of personality for use in the New Zealand context. This measure is known as the Mini-IPIP6 (Sibley et al., 2011). The scale extends the previous five-factor Mini-IPIP inventory initially developed by Donnellan, Oswald, Baird, and Lucas (2006). In this paper I present an analysis of the item discrimination and difficulty parameters for the Mini-IPIP6 using Item Response Theory (Samejima, 1969). As I outline below, unlike classical psychometric assessment, Item Response Theory examines the extent to which a set of items provide well-distributed measurement precision *across different levels of the latent trait they measure.* This study provides, for the first time, a detailed assessment of the response parameters for a public domain short-form measure of personality validated for use in New Zealand. To do so I analyse Mini-IPIP6 scores from the first wave of the New Zealand Attitudes and Values Study. This is a nationally representative longitudinal study of around 6000 New Zealanders.

What is Personality?

Previous research has typically identified five distinct factors, or broad clusters of related traits and behavioural tendencies, which constitute distinct latent dimensions of personality. These five broad-bandwidth dimensions of personality were synthesized and organized into a general framework by Goldberg (1981) who coined the term 'Big-Five' (see also Goldberg, 1990). This Big-Five model of personality contains the following factors: (1) Extraversion, (2) Agreeableness, (3) Conscientiousness, (4) Neuroticism, and (5) Openness to Experience. More recently, Ashton and Lee (2001, 2007, 2009) have made a compelling argument for an extended 'Big-Six' model of personality which adds an additional dimension to the mix: (6) Honesty-Humility. A descriptive summary of the core content and example traits for these different dimensions of personality is presented in Table 1.

Following from the general definition of personality by Mischel (2004) with which I began this manuscript; these six dimensions of personality reflect six distinct and 'highly meaningful patterns of variability in people's actions, thoughts, and feelings.' Why these six dimensions specifically? Evolutionary theory suggests that what we refer to as personality should reflect variation in behavioural systems or ways of acting that were equally adaptive for our species in different ecological niches (MacDonald, 1995, 1998; Nettle, 2006). Personality should reflect those traits in our species where it was sometimes the individuals high in the trait that had an adaptive advantage, but equally often in other situations, it was individuals low in the trait that had an adaptive advantage. Overall therefore, the traits had balanced selection pressures and this resulted in species-wide variation being maintained (Penke, Denissen, & Miller, 2007).

When we talk about personality, this is what we should be aiming to measure: traits which vary across people because such species-wide variation itself is the feature that has been selected for in evolution (Buss, 1991; Denissen & Penke, 2008). To give one example of the logic of defining personality as species-wide variation in behavioural systems

Table 1. Interpretation of each Mini-IPIP6 factor, including example traits, and likely adaptive benefit and costs resulting from high levels of each personality dimension. This table is taken from Table 3 of Ashton and Lee (2007, p. 156) with minor adaptations by Sibley et al. (2011) reinterpreting Neuroticism and Agreeableness within a Big-Five framework. (Ashton and Lee (2007) originally developed this framework for describing their HEXACO model of personality structure).

Factor	Interpretation	Example Traits	Likely adaptive benefits of high levels (in evolutionary history)	Likely costs of high level (in evolutionary history)
Extraversion	Engagement in social endeavours	Sociability, leadership, exhibition	Social gains (friends, mates, allies)	Energy and time; risks from social environment
Agreeableness	Ingroup co-operation and tolerance; reciprocal altruism in HEXACO model	Tolerance, forgiveness, (low) quarrelsomeness	Gains from cooperation, primarily with ingroup (mutual help and non-aggression)	Losses due to increased risk of exploitation in short-term exchange
Conscientiousness	Engagement in task-related endeavours	Diligence, organization, attention to detail	Material gains (improved use of resources), reduced risk	Energy and time; risks from social environment
Neuroticism (low Emotional Stability)	Monitoring of inclusionary status and attachment relations; kin altruism in HEXACO model.	Anxiety, insecurity, (low) calmness	Maintenance of attachment relations; survival of kin in HEXACO model	Loss of potential gains associated with risks to attachment relations.
Openness to Experience	Engagement in ideas-related endeavours	Curiosity, imaginativeness, (low) need for cognitive closure and (low) need for certainty	Material and social gains (resulting from discovery)	Energy and time; risks from social and natural environment
Honesty-Humility	Reciprocal altruism (fairness)	Fairness, sincerity, (low) entitlement and (low) narcissism	Gains from co-operation (mutual help and non-aggression)	Loss of potential gains that would result from the exploitation of others (and in particular outgroup members)

resulting from balanced selection pressure, Ashton and Lee (2007) argued that a high level of Conscientiousness should have been beneficial for our ancestors to the extent that it led to material gains and the improved use of available resources. However, a high level of Conscientiousness would also have caused the individual to expend time and energy in planning and organization, which would have come at the expense of other activities and may not always have been necessary in order to maximize gains.

A high level of Conscientiousness may have also led to increased social risks to the extent that others could free-ride or exploit outcomes resulting from planning and organization by the individual in question (e.g., food stores). In environments where expending energy in long-term planning and attention to detail were necessary to maximize gains, people high in Conscientiousness should have prospered. However, in environments where long-term planning did not yield any additional benefits, people low in Conscientiousness would have had an adaptive advantage because they would not have expended unnecessary energy or time on such endeavors and would have instead maximized immediate gains without longer-term associated costs. These trade-offs were presumably balanced across ecological niches (Penke et al., 2007), and this is why we see variation in this trait across people. We call this variation personality.

The Mini-IPIP6 Measure of Personality

There are a number of excellent (valid and reliable) public domain measures of personality available. These include (to name but a few), 50-item and 100-item instruments based on the International Personality Item Pool (IPIP; Goldberg, 1999), the 44-item Big Five Inventory (BFI; John & Srivastava, 1999), the 50-item Five Individual Reaction Norms Inventory (FIRNI; Denissen & Penke, 2008), the 100-item Big Five Aspects Scale (BFAS; De Young, Quilty & Peterson, 2007), and the 60- and 100-item HEXACO (Ashton & Lee, 2009). There are also a number of copyright personality inventories, the most well-known of which is possibly the NEO-PI-R (Costa & McCrae, 1992). I recommend the use of one of the many excellent public domain instruments (see Goldberg, 1999; Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006, for further discussion of the benefits of personality measures in the public domain).

Given their length, the various instruments listed above may not always be appropriate, however. In some research designs, where space is limited or there are constraints on interview time, a shorter measure of personality using a small select set of marker items for each personality dimension is needed. The Five-Factor Mini-IPIP is one such measure. The Mini-IPIP is a short-form public domain personality instrument initially developed by Donnellan et al. (2006) to assess the five broad-bandwidth dimensions of personality identified in the Big-Five framework (see also Gosling, Rentfrow & Swann, 2003, for an even shorter measure). One strength of this short-form measure is that the items were selected from the IPIP. The IPIP is a large-scale collaborative effort to develop a comprehensive system and set of items for personality measurement in the public domain (Goldberg, 1999).

Sibley et al. (2011) extended Donnellan et al's (2006) original Five-Factor Mini-IPIP to also include marker items for the sixth dimension of personality identified by Ashton and Lee (2007) in their Six-Factor (HEXACO) model of personality structure. Following Donnellan et al. (2011), Sibley et al. referred to this revised scale as the Mini-IPIP6. The Mini-IPIP6 is useful because it provides a way to briefly index the five dimensions of personality identified in the Five-Factor or Big-Five framework, *while also indexing* the sixth dimension of personality; reflecting HEXACO Honesty-Humility without altering the operationalization of the existing factors.

The six-factor HEXACO scale developed by Ashton and Lee (2009) is an excellent measure of personality. However, the HEXACO redefines many of the original Big-Five factors as rotational variants of their more traditional Big-Five counterparts, primarily Agreeableness and Neuroticism. This makes comparison across studies measuring Agreeableness within a Big-Five framework with studies assessing this dimension in a HEXACO framework quite tricky (see for example, Sibley, Harding, Perry, Asbrock, & Duckitt, 2010). The Mini-IPIP6, in contrast, retains Donnellan et al's (2006) short (four-item markers) of the original Five-Factor model and simply adds four marker items that load on a sixth rotated factor *without changing the existing structure*. The Mini-IPIP6 therefore provides a useful adaption in specific research 'niches' where one wants the balance of retaining markers within a five-factor personality model, but also the flexibility to index the additional Honesty-Humility personality dimension identified by Ashton and Lee (2007).

Sibley et al. (2011) validated the Mini-IPIP6 using Exploratory and Confirmatory Factor Analyses. In their original presentation of the Mini-IPIP6, they showed that the 24 items reliably fit a six-factor solution, with each four-item set loading on their hypothesized personality factor. These results provided good evidence for a six-factor model of personality indexed by the Mini-IPIP6. Sibley et al. (2011) also provided formal construct definitions for each of the six dimensions of personality, and these are presented in Table 1. In addition, Sibley et al. (2011) described a series of regression models showing that each of the Mini-IPIP6 dimensions predicted unique variance in concurrent criterion outcomes. For instance, the Mini-IPIP6 measure of Extraversion predicted how much time people spent socializing with their friends, whereas the Mini-IPIP6 measure of Consciousness predicted how much time people spent doing housework. These are exactly the type of outcomes that these different dimensions of personality should predict.

In sum, Sibley et al. (2011) provided good evidence validating the Mini-IPIP6. However, the psychometric analyses reported by Sibley et al. (2011) were based on a classical test theory framework (Exploratory and Confirmatory Factor Analysis) and do not tell us anything about the extent to which the Mini-IPIP6 items vary in their level of precision across the latent trait range. This is what Item Response

Theory allows us to assess.

Item Response Theory

Item Response Theory is a general method for modelling the precision (or reliability) of a set of items across different levels of a latent trait. For example, in education, an 'easy' test might reliably differentiate 'very poor' students from everyone else, but be less reliable in differentiating 'excellent' students from everyone else. Similarly, the Mini-IPIP6 measure of Extraversion might be better (i.e., more precise) at differentiating between people who are low in Extraversion from everyone else, relative to how accurate it is at differentiating between people high in Extraversion relative to others.

A reasonably even level of measurement precision is extremely important for a number of reasons. Skew in measurement precision means that a scale might be more reliable when measuring variability at the low level of a trait relative to variability at the high level of the trait. This can lead to biased estimates of the trait depending on a person's latent trait level. Such bias can also lead to inaccurate conclusions about the stability of the trait across time, as it might appear that people who are low in a trait change less in that trait over time, whereas people high in the trait may (spuriously) seem to change more in their trait level. Rather than reflecting genuine differential change, if measurement precision is uneven, this could simply be due to less reliable measures across time at a given trait level and hence more variability in the measure. This could make it look like people have changed more at one trait level relative to another.

So how does Item Response Theory actually work? To model the precision of a scale across the trait range, we need to know about two distinct parameters of each item. These are item difficulty and item discrimination. Stated formally, the logic behind a two-parameter logistic item response model (2PLM; Birnbaum, 1968) can be summarized as follows:

$$\textbf{(1.0)}\quad P_j(\theta_i) = 1 / (1 + \exp(-\alpha_j(\theta_i - \beta_j)))$$

This equation states that the probability that a given individual (*j*) with a given level of trait θ will have a level of that trait defined by one aspect of the person (their true trait level), and two aspects of the way it is measured (or item parameters). These two parameters are item difficulty (β_j) and item discrimination (α_j). In this model, trait levels can be thought of as reflecting a standardized (z-scored) range, with a Mean of 0 and Standard Deviation of 1.

Item difficulty reflects the level of the trait that a person would need to have a 1 in 2 (50%) chance of scoring in the positive direction on the item. For example, a person with the sample mean level of a trait ($\theta = 0$), would have a 50% chance of scoring in the positive (high trait) direction on an item with a difficulty value of 0. Similarly, a person with a trait level one unit above the mean ($\theta = 1$), would have a 50% chance of scoring in the positive (high trait) direction on an item with a difficulty value of 1.

What this means is that items that have higher difficulty values tend to be endorsed by fewer individuals (i.e., only those with higher levels of the trait). The term difficulty in this context arises from the fact that Item Response Theory tended to be used originally to model performance in educational assessments, where only students with a high latent academic ability would be likely to get a positive (correct) score on more difficult test items.

When examining ratings of Likert items, Item Response Theory provides a series of discrimination values in sequence for the set of (ordered) possible responses. That is, the lowest score, for example 1 versus any other score from 2-7; a score of 1 or 2 versus any other higher score from 3-7, and so on. With a 7-point Likert scale, there are therefore six item difficulty parameters, which reflect the following contrasts:

$$\textbf{(2.0)}\quad \begin{aligned} \beta_1 &= 1 \text{ v } 234567 \\ \beta_2 &= 12 \text{ v } 34567 \\ \beta_3 &= 123 \text{ v } 4567 \\ \beta_4 &= 1234 \text{ v } 567 \\ \beta_5 &= 12345 \text{ v } 67 \\ \beta_6 &= 123456 \text{ v } 7 \end{aligned}$$

Item discrimination, in contrast, reflects that ability of an item to differentiate between people with similar levels of the trait. Critically, an item's ability to differentiate between people is most precise at trait ranges corresponding to the item difficulty parameter. For example, imagine we have two items, one with a discrimination parameter of 1.0 and a difficulty of -1.0, the other also with a discrimination parameter of 1.0 but a difficulty parameter of 1.0. Both items are equally able to differentiate between individuals, but at different regions of the trait range.

The first item in this example would be better at differentiating between people with low levels of the trait, while the latter item would be better at differentiating between people with high levels of the trait. Conversely, the higher difficulty item would perform poorly when used to differentiate people at the low end of the trait range (people low on the trait are all fairly likely to get this 'hard' item 'wrong'), and the low difficulty item would perform poorly for differentiating between people at the high end of the trait range (people high on the trait would all be fairly likely to get this 'easy' item 'correct').

The difficulty and discrimination parameters can be combined to provide item Test Information Functions. By combining these functions, we can estimate the level of precision (i.e., reliability) of the entire scale across the entire trait range. You can get a good idea of how these parameters are combined to provide test information (I) by looking at the following equation:

$$\textbf{(3.0)}\quad I_j(\theta) = \alpha_j^2 \times P_j(\theta_i) \times (1 - P_j(\theta_i))$$

In this equation, α_j^2 is the squared item discrimination parameter for the *j*th item, and $P_j(\theta_i)$ is the probability of endorsing item *j* for individuals with a given (*i*) level of trait θ. A Test Information Function that looked like a bell curve centered on a score of $\theta = 0$ would indicate that the scale provided the most information about participants who were near the average level of the trait, but provided progressively less information about people at the high or low extremes of the trait range.

Item Response Theory thus provides information that is quite

distinct from that provided by classical test theory. Cronbach's alpha, for example, provides information on how well the items in a scale 'hang together' or inter-correlate, in the sense that they seem to be measuring the same thing. A Test Information Function, in contrast, provides information about the level of precision of a scale when assessing people with different levels of an underlying trait (see Hambleton & Jones, 1993, for discussion).

The desired shape of the Test Information Function, of course, depends upon the theoretical nature and expected prevalence of the trait in the population. For instance, in educational assessment, the ideal should be to develop tests that provide a high level of information across all levels of the trait range, say from 2 Standard Deviations below the mean, to 2 Standard Deviations above the mean, to give a very rough example. As such, we would hope that a 'good' test in this area would be relatively high and flat rather than bell-shaped. Similarly, research on Maori identity shows that the Houkamau and Sibley's (2010) multi-dimensional model of Maori identity provides reasonably well distributed test information across different levels of latent identification (Sibley & Houkamau, in press). This indicates that the measure provides a similar level of accuracy in differentiating between people with latent trait scores across a range of different levels of those scores (i.e., it reliably assesses both those highly identified, but also those with a low level of ethnic identity).

Well-designed personality instruments should also provide relatively high and flat Test Information Functions. That is, such measures should aim to differentiate between people equally across all levels of possible personality, rather than say, accurately differentiating between those low versus moderate or high on a trait, but being less accurate at differentiating between those moderate versus high. In contrast, the Test Information Function for a clinical measure of psychological health or distress should look quite different in a general population sample. Here we would reasonably expect that the Test Information Function would be skewed toward high values of θ, say for example, $\theta > 1$ (keeping in mind that 1 represents 1 Standard Deviation unit). A function of this type would indicate that the test provided detailed information differentiating between people with high versus very high levels of the trait in question, but did not differentiate that well between people with low or moderate scores. This is exactly the type of function observed, for instance, in Item Response Analyses of the Kessler-6 screening scale for non-specific psychological distress in general population samples (Kessler et al., 2002; Krynen, Osborne, Duck & Sibley, 2012).

Overview of the Present Study

The present study assessed the item response parameters of the Mini-IPIP6 in a nationally representative New Zealand sample. This is the second in a series of studies aiming to provide comprehensive psychometric information validating this public domain inventory for use in the New Zealand context. In the first study in this series my colleagues and I showed that the Mini-IPIP had a reliable six-factor structure with excellent convergent and discriminant validity (Sibley et al., 2011). Here I document item response parameters (discrimination and difficulty parameters as in Equations 1.0 and 2.0) and Test Information Functions for the four items that comprise each of the six Mini-IPIP6 subscales. This assesses the extent to which the scale reliably assesses the various dimensions of personality across different levels of the latent trait range. I expected that the Mini-IPIP6 would provide relatively even Test Information Functions distributed around the average estimated level of each latent trait, and extending ideally to +/- 2 Standard Deviation units.

Method

Sampling procedure

The NZAVS-2009 questionnaire was posted to 40,500 participants from the 2009 NZ electoral roll. The publicly available version of the roll contained 2,986,546 registered voters. This represented all citizens over 18 years of age who were eligible to vote regardless of whether or not they chose to vote, barring people who had their contact details removed due to specific case-by-case concerns about privacy. In sum, roughly 1.36% of all people registered to vote in New Zealand were contacted and invited to participate. The NZAVS-2009 sampled a total of 6,518 participants. The overall response rate (adjusting for address accuracy of the electoral roll and including anonymous responses) was 16.6%.

Participant details

Complete responses to all 24 Mini-IPIP6 items were provided by 5,576 participants (85% of the sample; 3298 women, 2278 men). Of those providing complete data, 72% were New Zealand European (n = 4,036), 16% of the sample were Māori (n = 915), 4% were of Pacific Nations ancestry (n = 222), 5% were of Asian ancestry (n = 254) and 3% were coded as 'other' (n=149). Participants' mean age was 47.02 (SD = 15.52). This is the same dataset which Sibley et al. (2011) analysed in their original factor analysis of the Mini-IPIP6.

Materials

Administration of the Mini-IPIP6 is described in Sibley et al. (2011). The 24 items in the scale were rated on a 7-point scale following the standard IPIP format developed by Goldberg (1999). This format asks participants to rate how well each statement describes them personally from 1 (very inaccurate) to 7 (very accurate). A copy of the Mini-IPIP6 inventory is presented in the Appendix.

Results

Overview of analytic strategy

I conducted a series of graded item response models examining response parameters for the items assessing each of the six Mini-IPIP6 factors separately. Analyses were conducted using Mplus version 6.11 with numerical integration (Muthén & Muthén, 2009). These analyses estimated item response parameters for each Mini-IPIP6 factor separately due to processing constraints, as Item Response Theory with numerical integration is computationally intensive by modern standards.

Item Response Analysis of the Mini-IPIP6

The item response models estimated two types of item parameter: an item

discrimination parameter (α) and a series of item difficulty parameters (β_1 - β_6) representing each set of ordered contrasts between different response options on the 7-point IPIP ordinal scale (see Equations 1.0 and 2.0). Discrimination and difficulty parameters for the Mini-IPIP6 are presented in Table 2.

As shown in Table 2, discrimination parameters for the Mini-IPIP6 items were all reasonably comparable, with values on or close to 1.0. This indicates that the Mini-IPIP6 items were all fairly comparable in providing similar levels of discrimination at their particular difficulty level. Difficulty parameters for the Mini-IPIP6 items were around -1.0 to -2.0 for the β_1 parameter at the low trait end, which represented the likelihood of responding in the negatively keyed direction (a low trait score of 1 versus 234567). Difficulty parameters for β_6, in contrast, were around 1.0 to 2.0 for most items. This parameter represented the likelihood of responding in the positively keyed direction (a high trait score of 7 versus 123456; see Equation 2.0). This suggests a reasonable spread of item difficulty across the trait range.

Test Information Functions for the six Mini-IPIP6 subscales are presented in Figure 1. These functions are based on the entire sample, and can thus be considered reasonably representative of the New Zealand population. These functions are graphed for θ values ranging from -3.0 to 3.0. This represents a broad range of values that should encompass the majority of the trait range in the New Zealand population.

As shown in Figure 1, the Test Information Functions for the Mini-IPIP6 indicated that the six personality measures provided the most information for θ values close to a score of 0. This indicates that the Mini-IPIP6 provided the most precise information about each latent trait for values close to the population mean. Moreover, the test information functions were reasonably flat for values of θ ranging from -1.0 to 1.0.

These results suggest that across the population as a whole, the Mini-IPIP6 seems to provide reasonably precise short-form measures of each of the six major broad-bandwidth dimensions of personality across a fairly broad range of each latent trait centered on average or mean levels of each trait. This is exactly what the Mini-IPIP6 is intended to provide, as it was developed as a general measure of personality that should be most precise in the normal trait range.

Discussion

In a recent article comparing various personality inventories, Grucza and Goldberg (2007) made the seemingly provocative statement that "Among the competing products developed by psychologists, perhaps the most important are their assessment instruments. Unfortunately, in psychology we have no Consumers Union to test competing claims and to compare these products on their overall effectiveness." (p. 167). I agree with this assessment and think it is important that we as a field continue to develop and evaluate freely available methods for assessing the constructs we seek to measure.

The purpose of the current manuscript was to apply recent advances in psychometric assessment to evaluate the measurement properties of a short-form personality inventory based on the IPIP format for use in the New Zealand context. This short-form measure, the Mini-IPIP6, is publicly available, and a copy is included in the appendix. The Mini-IPIP6 is based on the original Five-Factor Mini-IPIP developed by Donnellan et al. (2006), who in turn selected items from the IPIP developed by Goldberg (1999). The Mini-IPIP6 builds upon this earlier work by also including items that load on the distinct sixth 'Honesty-Humility' factor not indexed in earlier Five-Factor models. This is the second in a series of papers documenting the various properties and characteristics of the Mini-IPIP6 within the New Zealand population (see Sibley et al., 2011, for the first). In these papers, I hope to provide detailed and transparent information about the scale, its strengths, and its weaknesses, for the assessment of personality in the New Zealand context.

Results from a series of unidimensional graded response models indicated that the Mini-IPIP6 provides reasonably well distributed estimates of each of the six dimensions of personality across each latent trait range. Moreover, the Mini-IPIP6 scales were most precise when measuring levels of each personality trait that were close to the population average. This is entirely as expected, given that the scales were designed to assess variation in the typical trait range, rather than, in contrast, variation at the extremes of a trait range as might be the case for a measure of depression or clinical anxiety (see for example Krynen et al., 2012).

Recommendations for scale scoring

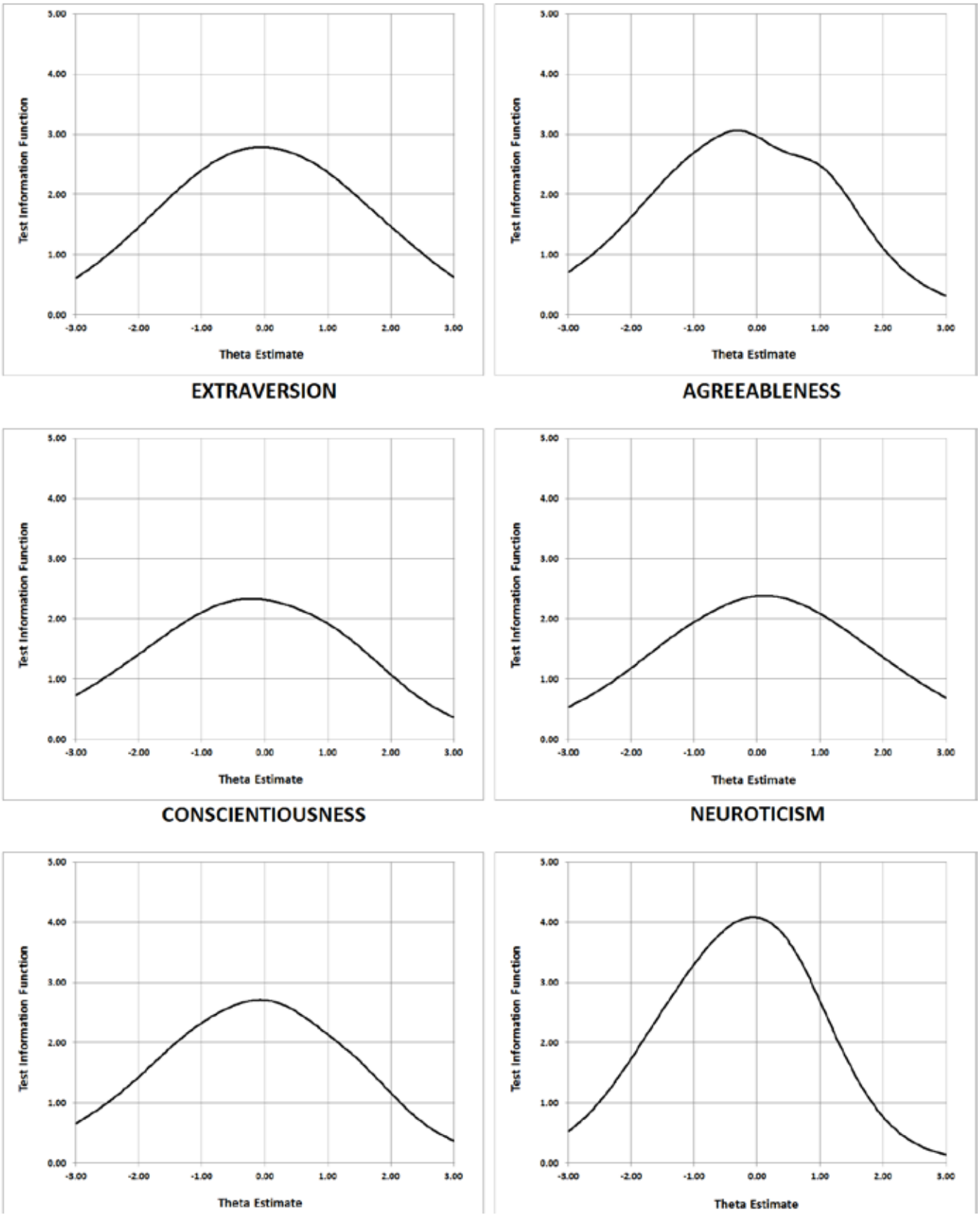
The Mini-IPIP6 can be scored using either a classical measurement model (by taking the average of scale items or estimating a latent variable in a structural equation model), or a more advanced IRT scoring method based on the parameters reported here. For the most part, the two scoring methods should generally yield similar results. For the majority of research on personality, Mini-IPIP6 scale scores can be calculated simply by first recoding the scale items worded in the opposing (low trait direction), and then taking the average score for the items in that subscale (i.e., summing the scores for the items in a given subscale, and then dividing that number by how many items there are in the subscale). This provides mean subscale scores, the method employed by Sibley et al. (2011) in their earlier work using the Mini-IPIP6. This scoring method should be appropriate for the majority of research focusing on assessing the extent to which different aspects of personality are linked to other outcomes of interest.

The difficulty and discrimination parameters reported in this paper could also be employed to score the Mini-IPIP6 using a more advanced IRT method. An IRT-weighted scoring procedure will be more reliable than simply creating a mean scale score as it is weighted based on item discrimination parameters and thus provides more reliable estimates for a given person depending upon their level of given personality trait. Those familiar with IRT could do this by applying the parameters reported here to scale people on the Mini-IPIP6 using one of the many available IRT scoring software packages. IRT-weighted Mini-IPIP6 scale scores will tend to be more precise at low and high levels of

Table 2. Discrimination (α) and difficulty (β_1, β_2, β_3, β_4, β_5, β_6) parameter estimates for the Mini-IPIP6 based on a series of graded response models.

	Item Response Parameters						
	α	β_1	β_2	β_3	β_4	β_5	β_6
Extraversion							
Am the life of the party.	1.07	-1.85	-1.04	-.21	.89	1.98	2.76
Don't talk a lot.	.84	-2.82	-1.67	-.80	.10	.86	1.91
Keep in the background.	1.00	-2.51	-1.32	-.49	.45	1.23	2.44
Talk to a lot of different people at parties.	.92	-2.25	-1.27	-.54	.24	.97	1.96
Agreeableness							
Sympathize with others' feelings.	1.46	-3.19	-2.51	-1.86	-1.19	-.28	.99
Am not interested in other people's problems.	.66	-3.74	-2.51	-1.59	-.76	.22	1.76
Feel others' emotions.	1.12	-3.15	-2.36	-1.70	-.92	.03	1.37
Am not really interested in others.	.81	-3.77	-2.69	-1.94	-1.19	-.28	1.25
Conscientiousness							
Get chores done right away.	.90	-3.39	-2.13	-1.18	-.27	.57	1.64
Like order.	.85	-3.49	-2.72	-2.02	-1.06	-.20	1.12
Make a mess of things.	.77	-4.21	-2.93	-2.05	-1.07	-.18	1.38
Often forget to put things back in their proper place.	.94	-2.63	-1.73	-1.17	-.64	-.09	1.11
Neuroticism							
Have frequent mood swings.	1.13	-1.32	-.23	.36	1.04	1.72	2.53
Am relaxed most of the time.	.77	-2.24	-.70	.38	1.48	2.57	3.92
Get upset easily.	.90	-2.15	-.76	.05	.89	1.72	2.80
Seldom feel blue.	.65	-2.82	-1.01	-.19	.76	1.80	3.15
Openness to Experience							
Have a vivid imagination.	.54	-4.22	-2.68	-1.52	-.21	.94	2.47
Have difficulty understanding abstract ideas.	1.10	-2.70	-1.72	-1.00	-.17	.47	1.61
Do not have a good imagination.	.79	-3.45	-2.35	-1.56	-.85	-.11	1.13
Am not interested in abstract ideas.	1.24	-2.57	-1.71	-1.12	-.29	.41	1.43
Honesty-Humility							
Feel entitled to more of everything.	.91	-3.43	-2.67	-1.89	-1.10	-.42	.71
Deserve more things in life.	1.17	-2.32	-1.69	-1.08	-.33	.17	.99
Would like to be seen driving around in a very expensive car.	1.47	-1.92	-1.42	-.97	-.52	-.16	.48
Would get a lot of pleasure from owning expensive luxury goods.	1.16	-2.08	-1.30	-.71	-.12	.31	1.10

Figure 1. Test Information Functions for the Mini-IPIP6 (Big-Six) Factors of Personality in a representative New Zealand sample (*n* = 5562).



personality, but should be reasonably comparable to mean scale scores for the majority of people who fall in the middle of the trait range. IRT-weighted scoring may be particularly important when one wants to maximize measurement precision in a research design, such as when the conclusions may have important real-world implications for social policy. I strongly recommend IRT-weighted scoring for research designs where the aim is to select people based on an extreme low or high trait level, such as might be the case in some specific instances of personnel selection.

Conclusion

These findings indicate that the Mini-IPIP6 provides a brief measure of personality that is reasonably well distributed in precision across the latent trait range for each of the six major dimensions of personality. Taken together with the initial validation study conducted by Sibley et al. (2011), the Mini-IPIP6 appears to provide a valid and reliable short-form measure of the six major dimensions of personality in the New Zealand context. Certainly, it is the inventory for which the most transparent and detailed validation information in New Zealand is currently publicly available. I hope that the presentation and validation of this short-form and easily administered instrument will provide a foundation for future research on personality in New Zealand. Moreover, given its brevity and satisfactory psychometric properties, I hope that these results will help other researchers to make informed decisions about which of the many available personality inventories to include in their research. In this regard, I hope that the Mini-IPIP6 will be deemed useful in other large-scale population surveys of the New Zealand population. This might help us progress toward a standard, validated, public domain format for assessing and comparing the effects of personality on various outcomes across diverse settings and research designs in the New Zealand context.

Acknowledgements

Collection of the New Zealand Attitudes and Values Study (NZAVS) data analysed in this paper was funded by The University of Auckland FRDF (# 3624435) and ECREA (# 3626075) grants awared to Chris Sibley.

References

Ashton, M.C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality, 15,* 327-353.

Ashton, M.C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150-166.

Ashton, M.C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91,* 340-345.

Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability.* In F. M. Lord & M. R. Novick, Statistical theories of mental test scores (pp. 397-472). Reading MA: Addison-Wesley.

Buss, D.M. (1991). Evolutionary personality psychology. *Annual Review of Psychology, 42,* 459-491.

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the five-factor model of personality: first steps towards a theory-based conceptual framework. *Journal of Research in Personality, 42,* 1285-1302.

DeYoung, C.G., Quilty, L.C., & Peterson, J.B. (2007). Between facets and domains: 10 aspects of the Big-Five. *Journal of Personality and Social Psychology, 93,* 880–896.

Donnellan, M.B., Oswald, F.L., Baird, B.M., & Lucas, R.E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18,* 192-203.

Costa, P. T., & McCrae, R. R. (1992). The NEO Personality Inventory (Revised). Odessa, FL: Psychological Assessment Resources.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 9,* 1216-1229.

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe,* Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84-96.

Gosling, S.D., Rentfrow, P. J., & Swann, W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37,* 504-528.

Grucza, R. A., & Goldberg, L. R. (2007). Comparative validity of 11 modern personality inventories: predictors of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment, 89,* 167-187.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices, 12,* 38-47.

Houkamau, C. A., & Sibley, C. G. (2010). The Multi-Dimensional Model of Māori Identity and Cultural Engagement. *New Zealand Journal of Psychology, 39,* 8-28.

John, O.P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: Guilford Press.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., … & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32,* 959-976.

Krynen, A., Osborne, D., Duck,

I. M., & Sibley, C. G., (2012). The Kessler-6 population screening measure of psychological distress: item response properties and ethnic group differences in New Zealand. *Manuscript submitted for publication.*

MacDonald, K. (1995). Evolution, the five-factor model, and levels of personality. *Journal of Personality, 63,* 525-567.

MacDonald, K. (1998). Evolution, culture, and the five-factor model. *Journal of Cross-Cultural Psychology, 29,* 119-149.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology, 55,* 1-22.

Muthén, L. K. & Muthén, B. O. (2009). *Mplus User's Guide* (5. ed.). Los Angeles, CA: Muthén & Muthén.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist, 61,* 622-631.

Penke, L., Denissen, J. J. A., & Miller, G. F. (2007). The evolutionary genetics of personality. *European Journal of Personality, 21,* 549-587.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

Sibley, C. G., Harding, J. F., Perry, R., Asbrock, F., & Duckitt, J. (2010). Personality and prejudice: Extension to the HEXACO personality model. *European Journal of Personality, 24,* 515-534.

Sibley, C. G., & Houkamau, C. A. (in press). The multi-dimensional model of Māori identity: Item Response Theory analysis of scale properties. *Cultural Diversity and Ethnic Minority Psychology.*

Sibley, C. G., Luyten, N., Purnomo, M., Moberly, A., Wootton, L. W., Hammond, M. D., … & Robertson, A. (2011). The Mini-IPIP6: Validation and extension of a short measure of the Big-Six factors of personality in New Zealand. *New Zealand Journal of Psychology, 40,* 142-159.

Corresponding Author:
Chris Sibley
School of Psychology
University of Auckland
Private Bag 92019
Auckland
c.sibley@auckland.ac.nz

© This material is copyright to the New Zealand Psychological Society. Publication does not necessarily reflect the views of the Society.

Appendix

The Mini-IPIP6

Instructions: **This part of the questionnaire measures your personality. Please circle the number that best represents how accurately each statement describes you.**

		Very Inaccurate						Very Accurate
I...		↓						↓
1.	Am the life of the party.	1	2	3	4	5	6	7
2.	Sympathize with others' feelings.	1	2	3	4	5	6	7
3.	Get chores done right away.	1	2	3	4	5	6	7
4.	Have frequent mood swings.	1	2	3	4	5	6	7
5.	Have a vivid imagination.	1	2	3	4	5	6	7
6.	Feel entitled to more of everything.	1	2	3	4	5	6	7
7.	Don't talk a lot.	1	2	3	4	5	6	7
8.	Am not interested in other people's problems.	1	2	3	4	5	6	7
9.	Have difficulty understanding abstract ideas.	1	2	3	4	5	6	7
10.	Like order.	1	2	3	4	5	6	7
11.	Make a mess of things.	1	2	3	4	5	6	7
12.	Deserve more things in life.	1	2	3	4	5	6	7
13.	Do not have a good imagination.	1	2	3	4	5	6	7
14.	Feel others' emotions.	1	2	3	4	5	6	7
15.	Am relaxed most of the time.	1	2	3	4	5	6	7
16.	Get upset easily.	1	2	3	4	5	6	7
17.	Seldom feel blue.	1	2	3	4	5	6	7
18.	Would like to be seen driving around in a very expensive car.	1	2	3	4	5	6	7
19.	Keep in the background.	1	2	3	4	5	6	7
20.	Am not really interested in others.	1	2	3	4	5	6	7
21.	Am not interested in abstract ideas.	1	2	3	4	5	6	7
22.	Often forget to put things back in their proper place.	1	2	3	4	5	6	7
23.	Talk to a lot of different people at parties.	1	2	3	4	5	6	7
24.	Would get a lot of pleasure from owning expensive luxury goods.	1	2	3	4	5	6	7

Scoring instructions. First, reverse code the following items: 6, 7, 8, 9, 11, 12, 13, 15, 17, 18, 19, 20, 21, 22, and 24. Next, create an average score for the four items assessing each dimension of personality. Extraversion: 1, 7, 19 and 23. Agreeableness: 2, 8, 14 and 20. Conscientiousness: 3, 10, 11 and 22. Neuroticism: 4, 15, 16 and 17. Openness to Experience: 5, 9, 13 and 21. Honesty-Humility: 6, 12, 18 and 24. An SPSS data entry template and scoring syntax is available from the author upon request.