

# FULL PAPER: Exploring Brand Personality Scale Development Using Rasch Modelling

*Francisco Conejo, University of Colorado - Denver, USA,  
[francisco.conejo@ucdenver.edu](mailto:francisco.conejo@ucdenver.edu)  
Ben Wooliscroft, University of Otago, New Zealand  
Andrea Insch, University of Otago, New Zealand*

**Abstract:** Scales in marketing rarely comply with measurement theory's unidimensionality, invariance and concatenation requirements. To address this, Rasch Modelling is applied to the Brand Personality (BP) construct, redefined as the set of human mental traits consistently associated to brands across situations and time. Ten Rasch BP scales are developed, positive and negative ones for each Big Five personality dimension. A first step towards actual BP measures, these scales lay the foundations for refinement. Addressing the notion of measurement itself, this paper highlights the importance of considering constructs from an intensity perspective, likely fertile ground for future marketing research.

**Keywords:** Marketing, brand, personality, definition, scale, Rasch, factor analysis

## Introduction

Scales purporting to measure Brand Personality (BP) have been advanced since the 1950s. The most influential, by far, is that of Aaker (1997). Impressed by its rigorous factor-analytical methodology, and reported validity, reliability and generalisability, the marketing community largely embraced it as the definitive BP instrument. Whereas other BP efforts have had negligible impact, Aaker can be credited for having started an important and growing research stream.

However, Aaker's (1997) BP instrument faces growing concern. It is on the one hand questioned in regards to validity. A significant portion of its items are outside the realm of personality as understood by psychology (Azoulay and Kapferer 2003; Bosnjak, Bochmann, and Hufschmidt 2007). On the other hand, the instrument's methodology has also come into question. Austin et al. (2003) requested its original data set to empirically address a series of issues, however, this was impossible as "*the Aaker data are no longer in existence as the result of a massive computer failure at Stanford University*" (p. 91).

Instead of developing yet another factor-analytical BP taxonomy, the present paper explores the application of an alternative scaling technique, Rasch Modelling (RM), to BP. This follows calls that more rigorous BP scales be developed (Austin et al. 2003; Azoulay and Kapferer 2003), particularly through unconventional approaches (Romaniuk 2008), whose discontinuous nature is often necessary for fields to advance (Kuhn 1962; Ladik and Stewart 2008). Exploratory in nature, the present research makes inroads towards fully-operational BP scales. It also contributes towards marketing's scaling methodology, suggesting RM as a viable technique for other constructs. More generally, this paper raises awareness of several important, albeit neglected issues. It thus contributes towards the discussion of what measurement is, or ought to be, within marketing.

*The authors would like to thank MB editors and reviewers for their insightful and encouraging comments.*

## Measurement in Marketing

As a mature science marketing has become rather quantitative. Inspection of any given journal reveals how constructs, features and interrelationships are constantly 'measured'. However, the notion of measurement itself is oddly neglected by the literature. This might seem unnecessary given its self-evident nature. But as Low and MacMillan (1988) suggest, it is sometimes healthy for fields to pause, evaluate their work, and identify new directions.

Marketing typically equates measurement with statistical analysis. This means that marketing, as with business generally, understands measurement as the assignment of numbers to the characteristics of objects according to specified rules, for example, Weiers (1988), Tull and Hawkins (1990) or Aaker D *et al.* (2005). This notion, popularised by Stevens (1946), seems scientific. Yet it neglects a fundamental issue: not all number-assignment rules lead to proper measures. This vital condition is merely presumed (Michell 1997), and is a product of confusing statistics with measurement (Gaito 1980).

Within measurement theory (MT), number assignment refers to coding. Scale development goes well beyond this, as it derives quantitative latent variables from coded observations (Salzberger and Koller 2013). Proper measures, as understood by MT and used in physical sciences, have standardised quantities against which other magnitudes are compared (Michell 1997). Proper measurement units thus comply with three fundamental criteria: They must be *unidimensional*, referring to single constructs at a time; *invariant*, of consistent magnitude irrespective of respondent, situation, and objects measured; and *concatenable*, addible with one another to express different amounts (Andrich 1988). Proper measure development therefore goes well beyond mere number assignment, which has little to do with the notion of quantity (Salzberger and Koller 2013).

Furthermore, marketing scales are largely based on raw scores. However, these lack intrinsic meaning (Andrich 1988). As Churchill (1979, p. 72) admits, raw scores are "... *not particularly informative about the position of a given object on the characteristic being measured because the units in which the scale is expressed are unfamiliar.*" Without units there is no measurement (Michell 1997), and without units results cannot be properly discussed, validated, and synthesised into coherent bodies of knowledge. If a field is to truly develop as a science, measurement units become essential (Salzberger and Koller 2013).

Stevens (1946) admits that his lenient number-assignment measurement notion, proposed given psychology's inability to meet MT cannons, is arbitrary and mathematically incorrect. Not constituting proper measurement, he cautions against it. However, he defends this *illegal statisticing* through the useful results it might nevertheless produce. This attitude may be seen across social science, including business, where in regards to measurement pragmatism and tradition prevail (Michell 1997).

An example of the above is Factor Analysis (FA), the technique through which most BP instruments have been developed, including Aaker's (1997). While efforts might be psychometrically robust, the technique poses limitations from a MT perspective (Smith AM 1999; Waugh and Chapman 2005). First and foremost, FA does not produce measures (Bond and Fox 2015; Wright 1988, 1994). It instead uncovers hierarchical item structures (Goldberg and Digman

1994; Linacre 2009c). This follows FA's original purpose, to organise and reduce data into manageable taxonomies (Smith RM 1996; Thurstone 1934). Second, factors lack intrinsic order. They are but collections of statistically-related items (DeVellis 2012). Closeness to latent variables is indeed indicated. Though items' location along the latent variables, essentially their measure, remains unanswered (Schumacker and Linacre 1996). Third, factors are statistical, and thus context-dependent. They are applicable only to situations from which they were derived. They vary greatly with respect to instruments, stimuli and samples used (Bond and Fox 2015). Finally, factors may be misleading (Steinberg and Thissen 1996). Items loading under a factor, while statistically related, are not necessarily conceptually related (Linacre 2009b). Conversely, thematically-related items may end up scattered across different factors (Wright 1988). This leads factors to acquire different meanings, hence names, as evidenced by the plethora of inconsistent *BP* scales developed to date.

Thurstone (1974), one of FA's developers, cautions against the technique for scaling purposes. Yet MT is often neglected in social science (DeVellis 2012). Pragmatism overshadows theory, and factors are deemed *close enough* to proper measures (Michell 1997). As Cronbach (of  $\alpha$  fame) (1951, p. 297) admits, "*In designing tests... Scalability is not a requisite.*" Kaiser (of KMO fame) (1970, p. 403) adds, as to whether FA produces scales, "*Rather than fuss and fight about it, avoid the question...*".

Salzberger (2009) reviews different scale development approaches. He concludes that, from a MT perspective, conventional techniques pose limitations. Though he notes that Rasch Modelling is quite adequate, and able to deliver close approximations to proper measures as understood by MT (Andrich 1985; Salzberger and Koller 2013). This paper thus explores Rasch-based scaling towards the development of *BP* measures that more closely approximate MT guidelines.

## Rasch Model Background

Rasch Modelling (RM) is named after its developer, Danish mathematician Georg Rasch (1901-1980). The technique was developed during the 1950s (Rasch 1960/1993), though builds on the item-response school of measurement dating back to the early 20<sup>th</sup> century (Hambleton and Swaminathan 1985). RM's rigorous criteria strive to meet MT requirements and produce closer proper measure approximations (Andrich 1985; Linacre 2009b). It does so by avoiding the correlational approach of techniques such as factor analysis. It instead applies a multiplicative Poisson algorithm to establish probabilistic relationships between items' intensity and respondents' propensity in regards to a latent variable (Wright 1999). This transforms nominal or ordinal scores, common in social science, into logarithms (Andrich 1988). These new continuous units allow items and respondents to be placed along the latent variable's intensity continuum, that is, to be measured (Meads and Bentall 2008). Mathematically, RM's formula is:

$$(1) \quad P_{ni} = e^{(B_n - D_i)} / (1 + e^{(B_n - D_i)}), \text{ where}$$

$P_{ni}$  is the probability  $P$  of respondent  $n$  endorsing item  $i$ ;  $e$  the natural logarithmic constant 2.718;  $B_n$  the propensity of respondent  $n$ ; and  $D_i$  the intensity of item  $i$ .

Despite being called *the* Rasch Model, there is actually a whole family of models (Andrich 1988). The dichotomous model, above, has been extended to cover different testing situations like rating scales (Andrich 1978), partial credit responses (Masters 1982), and faceted testing (Linacre 1992). More complex, these extensions are still heavily based on the original model (Linacre 2009b). Technical discussions on the different Rasch models may be found in Andrich (1988) or Salzberger (2009), among others.

RM has been extensively used across social science (Bond and Fox 2015). The technique is gaining attention in marketing, with its versatility making it viable for diverse contexts. Rasch marketing efforts include, among others, Soutar et al.'s (1990) durable good acquisition order; Laurent, Kapferer et al.'s (1995) brand awareness structure; Salzberger's (2000) analysis of Shimp and Sharma's (1987) consumer ethnocentrism scale; Ganglmair-Wooliscroft and Lawson's (2003) Affective Response to Consumption (ARC) scale; De Bastiti et al.'s (2005) service quality scale; Ewing et al.'s (2005) cross-cultural advertising assessment; Ganglmair-Wooliscroft and Wooliscroft's (2010) consumer language assessment; Ganglmair-Wooliscroft's (2010) durable good ownership patterns; Ganglmair-Wooliscroft and Wooliscroft's (2013) cross-cultural ARC assessment; Salzberger et al.'s (2014) consumer gender bias assessment; and Wooliscroft et al.'s (2014) ethical consumption hierarchy. Within literature on BP, no empirical efforts have used RM.

## Methodology

RM is superior to conventional quantification in several ways (Nijsten, Unaeze, and Stern 2006; Salzberger and Koller 2013). However, its purpose is not to outright replace mainstream approaches. The technique should instead be seen as a complement, to further improve measurement standards (Andrich 1988; Salzberger 1999). To build on familiar territory, scale development began by following the first three stages of Churchill's (1979) framework.

## Domain Specification

Theoretical conceptualisations should be literature-based (Churchill 1979). Yet despite decades of research, there is still no commonly-accepted BP definition (Smit, Van Den Berge, and Franzen 2003). The literature is replete with inconsistent, even questionable notions. Most notable is Aaker (1997, p. 347), who defines BP as "... *the set of human characteristics associated to a brand.*" This broad, ambiguous definition refers to human features in general, not only mental traits. It incorporates physical attributes (for example, *intelligent, good looking, rugged*); demographics (for example, *young, small-town, upper class*); and social evaluations (for example, *wholesome, successful, corporate*). All these are outside the realm of human personality (Allport 1938; Azoulay and Kapferer 2003; Bosnjak et al. 2007). Aaker's (1997) taxonomy might refer to a personified brand image, but not to brand personality in a strict psychological sense. Neither do most other BP instruments, which follow or slightly modify Aaker's notion.

Churchill (1979) suggests that if conceptual clarity is impossible, new definitions should be advanced. BP derives from human personality (Smit et al. 2003), so conceived properly, BP is its strict equivalent (Bosnjak et al. 2007). BP is thus re-defined as *the set of human mental traits consistently associated to brands across situations and time*. This proposed BP definition matches psychology's general personality notion, conceived as an individual's innate, pervasive and enduring mental characteristics, which lead to distinct patterns of behavior consistent across

situations and time (Allport 1938; Cervone and Pervin 2008). The proposed definition also differentiates BP from other branding constructs, which is key if terminological order is to be brought to the field (Conejo and Wooliscroft 2015).

### Operationalisation

Construct transformation from theoretical to measurable should also be literature-based (Churchill 1979). However, given improper domain specifications, extant BP instruments frequently contain items outside the realm of personality as understood by psychology. BP was therefore operationalised de novo, through personality traits taken directly from psychology using Goldberg's (1992) 100 Markers, shown in Table 1, below.

**Table 1: Goldberg's (1992) 100 Personality Markers by Big Five Dimension**

<b>Surgency</b>	<b>Agreeableness</b>	<b>Conscientiousness</b>	<b>Emotional Stab.</b>	<b>Intellect</b>
<b>Positive</b>	<b>Positive</b>	<b>Positive</b>	<b>Positive</b>	<b>Positive</b>
Extraverted	Kind	Organized	Unenvious	Intellectual
Talkative	Cooperative	Systematic	Unemotional	Creative
Assertive	Sympathetic	Thorough	Relaxed	Complex
Verbal	Warm	Practical	Imperturbable	Imaginative
Energetic	Trustful	Neat	Unexcitable	Bright
Bold	Considerate	Efficient	Undemanding	Philosophical
Active	Pleasant	Careful	<b>Negative</b>	Artistic
Daring	Agreeable	Steady	Anxious	Deep
Vigorous	Helpful	Conscientious	Moody	Innovative
Unrestrained	Generous	Prompt	Temperamental	Introspective
<b>Negative</b>	<b>Negative</b>	<b>Negative</b>	Envious	<b>Negative</b>
Introverted	Cold	Disorganized	Emotional	Unintellectual
Shy	Unkind	Careless	Irritable	Unintelligent
Quiet	Unsympathetic	Unsystematic	Fretful	Unimaginative
Reserved	Distrustful	Inefficient	Jealous	Uncreative
Untalkative	Harsh	Undependable	Touchy	Simple
Inhibited	Demanding	Impractical	Nervous	Unsophisticated
Withdrawn	Rude	Negligent	Insecure	Unreflective
Timid	Selfish	Inconsistent	Fearful	Imperceptive
Bashful	Uncooperative	Haphazard	Self-pitying	Uninquisitive
Unadventurous	Uncharitable	Sloppy	High-strung	Shallow

Several reasons support the above choice: Goldberg's set was purposely derived to operationalise the Big Five personality factors and facilitate subsequent research (Saucier 1994); unlike most BP instruments it offers both positive and negative traits; the set is reasonably comprehensive yet practical; and its items are easily understood.

### Data Collection

After promising pilot results, data was collected via convenience sampling around Dunedin, New Zealand. Participants filled out self-administered questionnaires. Respondents indicated which of the 100 personality traits applied to each of two brands. Items largely consisted of checklists for being efficient (Romaniuk 2008) and providing the dichotomous data required for RM. Polytomous/Likert answer options would have provided more detailed information. However,



dichotomous ones were preferred given the exploratory nature of the present effort, best simplified as much as possible. Dichotomous answer options were also preferred to reduce response time and effort given the 200+ items involved.

Traits were randomised and order-inverted to reduce biases. Responses were anonymous, though age, gender and brand usage were requested for analysis purposes. Surveys were easily completed in 20 minutes, normal within personality research (Brody 1994). To reduce target homogeneity (see Peabody and Goldberg 1989) stimuli were refocused from multiple top brands to a single polarised category. Apple and Microsoft were used for being well-known; dissimilar (symbolic vs. utilitarian (see De Chernatony and McWilliam 1990)); and eliciting intense feelings (Belk and Tumbat 2005). To benefit from reference frames (see Murphy, Moscardo, and Benckendorff 2007), brands were evaluated concurrently, checklists next to each another. To counter primacy or recency effects brands were rotated. To evoke richer associations brand logos were placed above checklists.

Collection produced 418 initial responses. 86 were eliminated for being anomalous: defective or duplicate surveys; incomplete, ambiguous or patterned answers; or visibly altered respondents. This left 332 usable responses, above Bond and Fox's (2015) 300 necessary upper limit. As to representativeness, a general, non-extreme sample suffices for RM (Andrich 1988). Compared to that used by, for example, Meads and Bentall (2008) to develop their Rasch Hypomanic Personality Scale, the present sample is more balanced in terms of composition, gender and age, thus deemed adequate (Table 2, below).

**Table 2: Present Sample Characteristics**

Sample	Size	Nature	% Female	Age Range	Age Ave.	Age Std. Dev.
M & B (2008)	318	Undergrad Only	68.9	18-48	21.5	4.00
<b>Present Sample</b>	<b>332</b>	<b>UG &amp; Gen. Public</b>	<b>51.5</b>	<b>17-67</b>	<b>26.9</b>	<b>10.97</b>

## Measure Development

Instead of FA, as suggested by Churchill (1979), RM was applied. The scale development sequence, based on the Rasch literature and software used, Winsteps v.3.69, consisted of two general phases: 1) Dimensionality Assessment, and 2) Measure Refinement. Unlike conventional approaches, the latter phase improves how well data fit the a-priori theoretical Rasch model (Bond and Fox 2015). This is accomplished through a series of steps: 2.1) Construct Map, 2.2) Differential Test and Item Functioning, 2.3) Item Polarity, 2.4) Item Fit, 2.5) Individual Item Functioning, and 2.6) Concurrent Item Functioning. These steps culminate with the Item Person Map, which illustrates refined scales. Since the development sequence for all BP dimensions was identical, and results fairly consistent, only *Surgency's* analysis is reported. General results are subsequently discussed.

### *1 Dimensionality Assessment*

Measures must refer to single constructs (Salzberger et al. 2014), however, personality is multi-dimensional, encompassing different constructs (Cervone and Pervin 2008). Data was thus

separated into five matrices, one for each Big Five dimension, to be analysed in isolation. Unidimensionality was nevertheless verified (Smith RM 1996).

Variance was first assessed. Within RM, modelled variance should be maximal, ideally over 75%. The remaining unexplained variance should conversely be minimal (Linacre 2014). *Surgency's* modelled variance was 26.1%, and unexplained variance a staggering 73.9%. This suggested something fundamentally wrong with *Surgency* as factor-analytically derived. Instead of referring to a single construct, the dimension seemed to comprise multiple constructs (Linacre 2009b).

A Rasch Principal Component Analysis (R-PCA) of item residuals was thus conducted. Unidimensional data has uncorrelated random residuals, while multi-dimensional data has *contrasts*, correlated residual clusters (Linacre 2014). R-PCA identified five *Surgency* contrasts: The first accounted for 17.8% of the unexplained variance, and the remaining ones accounted for 5.4 %, 4.7%, 4.3% and 3.9%. Smith et al. (2006) suggest that sub dimensions contain at least three items and that less are likely spurious. Contrast 1 contained 4.8 items, and the remaining contrasts 1.5, 1.3, 1.2 and 1.1. This confirmed a single sub-dimension, with *Surgency* composed of two independent, albeit related constructs.

Figure 1: Surgency Contrast 1 Residual Plot & Loadings

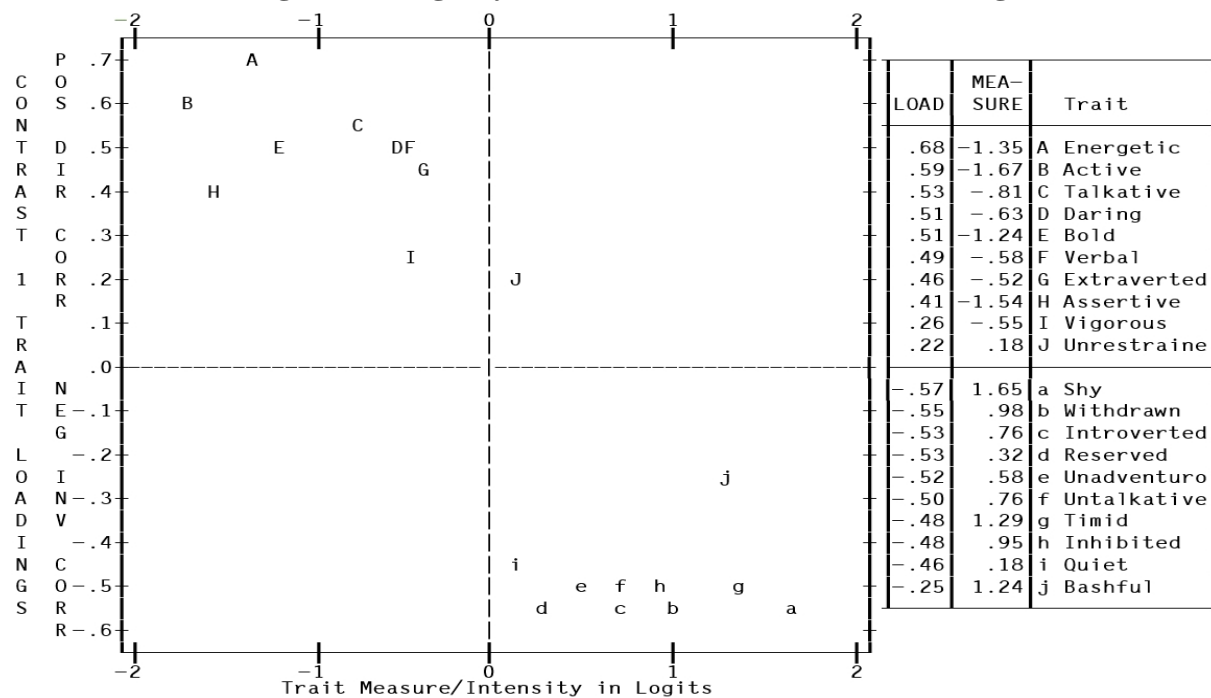


Figure 1, above, shows two distinct *Surgency* clusters: The first, upper-left, consists of positive traits only, for example, *extraverted*. Positive loads indicate direct correlation with *Surgency*, these traits making up the construct. The second cluster, lower-right, consists of negative traits only, for example, *introverted*. Negative loads indicate an inverse relation towards *Surgency*. These negative traits are not part of the *Surgency* construct, instead forming a separate sub-dimension (Linacre 2014).

This positive-negative trait clustering is consistent with Yamaguchi's (1997) findings. However, it was nevertheless verified. This was done by revisiting the original data set, splitting dimensions into their positive and negative sub-dimensions, and correlating subtotal respondent scores. In agreement with the above item loads, split dimensions consistently showed moderate negative correlations between positive and negative sub-dimensions. *Surgency's* correlated at -0.328; *Agreeableness'* at -0.581; *Conscientiousness'* at -0.570; and *Intellect's* at -0.483. The only exception was *Emotional Stability*, whose positive and negative sub-dimensions showed a near zero correlation of 0.037. This might be explained by the mismatch in item numbers: Contrary to the four other dimensions, each with 10 positive and 10 negative traits, *Emotional Stability* had only six positive traits and 14 negative traits. Regardless, the lack of significant positive correlations between positive and negative sub-dimensions further suggested treating them as separate variables.

Differently-valenced sub-dimensions, addressing distinct constructs, are better scaled individually than combined (DeVellis 2012; Linacre 2009b). *Surgency* was thus split into two separate sub-dimensions: *SurgencyP*, referring to outbound energy, containing positive traits (*active, assertive, bold, daring, energetic, extraverted, talkative, unrestrained, verbal* and *vigorous*); and *SurgencyN*, referring to inbound energy, containing negative traits (*bashful, inhibited, introverted, quiet, reserved, shy, timid, unadventurous, untalkative* and *withdrawn*).

### 1.1 SurgencyP Dimensionality

Personality consists of successively-specific sub-dimensions (Cervone and Pervin 2008). *SurgencyP* and *SurgencyN* were therefore inspected. Variance, contrasts and residual plots suggested that each might contain two further sub dimensions: *SurgencyP* showed *Extraversion* items (*talkative, assertive*), and *Activity* items (*energetic, vigorous*). *SurgencyN* showed *Withdrawal* items (*shy, inhibited*), and *Reticence* items (*reserved, untalkative*). These sub-dimensions are consistent with personality's literature (Cattell 1946; Deater-Deckard *et al.* 2009; Jang, Livesley, and Vernon 1996). Separating these sub-dimensions would have better complied with MT's unidimensionality requirement. However, *SurgencyP* and *SurgencyN* were kept intact for lack of empirical support: Variance, contrasts and residual plots were not conclusive enough to warrant further dimensional splits.

## 2. Measure Refinement

Having a preliminary *SurgencyP* measure, refinement was undertaken. RM is a theoretical ideal from which empirical data inevitably departs (Wright and Linacre 1994). Refinement encompasses a series of tests and items not adhering to model parameters are iteratively culled to improve compliance until an acceptable item set remains.

### 2.1 Construct Map

Before refinement commenced a Construct Map was developed. Items were a priori intensity-ordered to better understand the latent variable's progressive nature and help guide refinement (Linacre 2009b). To improve objectivity, intended trait meanings were looked up in the online Oxford English Dictionary. A basic dictionary was purposely used to reflect respondents' average English level. *SurgencyP's* Construct Map, Table 3, below, confirmed the *Extraversion* and *Activity* sub-dimensions, and a possible third one: *Assertiveness*. The exercise also revealed some items' redundancy like *daring/bold* or *vigorous/energetic*. Their meanings quite close, even synonymous, redundant pairs were earmarked for elimination.



**Table 3: SurgencyP Construct Map**

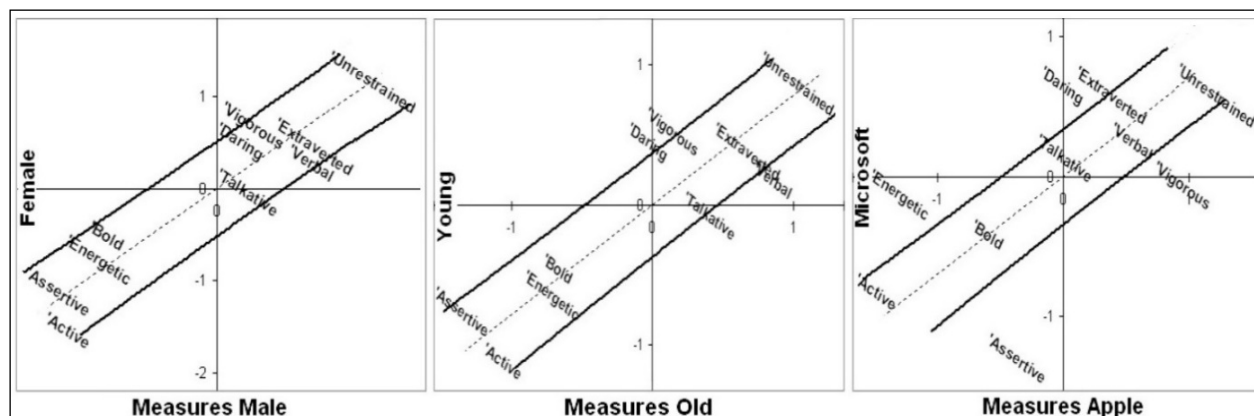
Sub-Dimension	Trait	Oxford English Dictionary (2016) Definition
Assertiveness	<b>Unrestrained</b>	Not restricted, reserved, moderated; Emotional, passionate, impulsive.
	<b>Bold</b>	Willing to take risks; Confident and courageous; Audacious, daring.
	<b>Daring</b>	Adventurous, audacious, bold.
	<b>Assertive</b>	Confident and forceful.
Activity	<b>Vigorous</b>	Strong, healthy, full of energy; Involving physical effort, forceful.
	<b>Energetic</b>	Showing great activity or vitality.
	<b>Active</b>	Operational, moving about; Alert and lively; Participating, involved.
Extraversion	<b>Extraverted</b>	Outgoing, socially confident.
	<b>Talkative</b>	Fond of or given to talking.
	<b>Verbal</b>	Relating to or in the form of words; Spoken rather than written, oral.

2.2 Differential Test and Item Functioning

MT requires that scales function consistently. Measures, that is, item locations along variables' intensity continua, need to be similar for different stimuli and respondent subgroups (Salzberger et al. 2014). Differential Test Functioning (DTF) determines whether the entire test, all items simultaneously, functions consistently (Linacre 2009b). *SurgencyP*'s data was thus split according to respondent gender, age and brand usage. Sub-sample pairs (male vs. female, younger vs. older, and Microsoft vs. Apple) were then compared, their item intensities ideally along the unitary diagonal, Figure 2, below.

*SurgencyP*'s three subsamples arranged items in the same general order. However, within each sub-sample pair about half the items were outside the 95% confidence interval. Those items function too unequally to contribute towards proper measures and must be eliminated (Linacre 2009c). Given the few initial items, only a single one, *verbal*, would have remained. This would have made it impossible to further process *SurgencyP*.

**Figure 2: SurgencyP DTF by Gender, Age & Brand Usage**



To obtain a partial result, and gain insights for future scaling efforts, a more lenient approach was adopted: Instead of eliminating traits exceeding confidence intervals just once, the three analyses were cross-referenced. Traits not complying with two of the three instances were deleted. This was

done iteratively. Non-complying items were eliminated one at a time until consistent item operation was achieved (Linacre 2009b). *Assertive, energetic, daring, vigorous, and unrestrained* were removed.

A Differential Item Functioning (DIF) test verified remaining items' invariance. Unlike DTF, which evaluates all test items concurrently, DIF assesses items one at a time while holding all others constant. This more precisely indicates the existence and magnitude of any item-respondent group/stimuli interactions that might distort measure performance (Linacre 2009b). *SurgencyP*'s DIF was acceptable. Sub samples arranged items in the same general order, and trait intensities for each were reasonably close. Only in one out of 15 instances did a trait, *active*, slightly exceed Zwick et al.'s (1999) 0.64 logit intensity difference guideline. Since no traits showed two DIF instances, all were retained.

### 2.3 Item Polarity

RM requires that respondent trait allocation propensity correlate with trait intensity. Point Measure Correlations (PMC) indicate to what extent these align. PMC values range from -1 to +1. While critical values fluctuate, PMC should be noticeably positive, over 0.65, to indicate strong correlation. Near zero and low PMC indicate a weak model fit. Negative correlations contradict latent variables' direction (Linacre 2009b). For *SurgencyP*, PMC were all noticeably positive: 0.63 (*active*), 0.65 (*bold*), 0.68 (*extraverted*), and 0.72 (*verbal and talkative*) (see Table 4). These suggest that trait intensity and respondent propensity align, adhering reasonably well to Rasch model parameters. While sound PCMs are necessary, they are not sufficient to indicate model compliance. Other indicators were also considered.

### 2.4 Item Fit

Fit further indicates how well data conforms to Rasch model parameters (Wright 1999). Different chi-square statistics might be applied for this (Linacre 2003). The one used by Winsteps is mean-squares (MNSQ), the average value of squared residuals (Bond and Fox 2015). MNSQ range from zero to infinity with an ideal value of 1. As MNSQ exceed 1, error increasingly hinders measure development, while as MNSQ fall below 1, stochasticity decreases (Wright and Linacre 1994). While MNSQ should approximate 1, no single critical interval exists for them (Smith RM, Schumacker, and Busch 1995). Guidelines vary according to test, item and respondent characteristics, but for general purposes a 0.5-1.5 interval suffices. As test importance increases, limits tighten around the 1.0 ideal. Given *SurgencyP*'s somewhat advanced refinement, and in line with high-stakes situations, the MNSQ fit interval was set at 0.8-1.2 (Wright and Linacre 1994).

MNSQ refer to fit magnitude. Though fit significance is also important. Each MNSQ thus has a corresponding standardized Z statistic, ZSTD. These show the probability of MNSQ as a unit normal deviate. ZSTD correspond to the null hypothesis of empirical data fitting the model. A ZSTD value of 0 is ideal. When ZSTD exceed +/- 1.96,  $p < 0.05$ , significance is sufficient to reject the null hypothesis. Statistically significant model misfit thus occurs when ZSTD equal or exceed +/- 2.0 (Linacre 2009b).

Two types of fit statistics are addressed by MNSQ and ZSTD. *Infit* is the weighted average of squared residuals. It gives relatively more importance to well-targeted respondents/items (Smith RM et al. 1995). Conversely, *outfit* is the unweighted average of squared residuals. It gives equal

importance to all respondents (Bond and Fox 2015). To capitalize on their respective strengths, both infit and outfit were used to assess model fit.

*SurgencyP*'s infit and outfit was generally acceptable, as shown in Table 4, below. All five traits were within the 0.8-1.2 MNSQ target. The only concern was *active*, whose outfit MNSQ of 1.19 is borderline. As to ZSTD, two traits were slightly beyond the +/- 2.0 target interval: *talkative* (-2.2 infit/-2.1 outfit), and *verbal* (-2.4 infit/-1.9 outfit). However, they did not warrant deletion being instances of over-fit, adhering too well to the model (Linacre 2009b).

**Table 4: SurgencyP PMC and Fit**

ENTRY NUMBER	MOD MEAS	MOD S.E	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Trait
			MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
5	-1.13	.12	1.11	1.8	1.19	1.7	.63	.66	74.0	77.8	Active
4	-.44	.11	1.07	1.4	1.14	1.9	.65	.68	73.1	72.4	Bold
1	.72	.11	1.03	.7	1.00	.0	.68	.69	69.3	71.9	Extravert
2	.24	.11	.90	-2.2	.88	-2.1	.72	.69	74.0	71.4	Talkative
3	.61	.11	.89	-2.4	.88	-1.9	.72	.69	75.3	71.2	Verbal
MEAN	.00	.11	1.00	-.1	1.02	-.1			73.1	72.9	
S.D.	.70	.00	.09	1.8	.13	1.7			2.1	2.5	

Respondent: REAL SEP.: .00 REL.: .00 ... Trait: REAL SEP.: 6.02 REL.: .97

### 2.5 Individual Item Functioning

RM produces probabilistic functions. These indicate how likely respondents allocate personality traits to brands. The graphic representation of these functions, Figure 3 below, are known as Item Characteristic Curves (ICC) (Bond and Fox 2015).

The horizontal axis, in logits, represents the difference between respondent item allocation likelihood and item intensity. The vertical axis, in percentages, represents the item allocation likelihood. The red logarithmic curve going from lower left to upper right represents the theoretical ICC. As one moves from left to right along the curve, odds increase in favour of item allocation. The thin green lines above and below theoretical ICCs represent 95% confidence intervals. These are 1.96 vertical standard deviations away from theoretical ICCs. The thick blue jagged lines represents items' empirical ICC. Each 'x' represents response categories' average value (Linacre 2009b). There being six response categories follows *SurgencyP* having five BP items. Respondents can attribute one to five items (five categories) or none at all (sixth).

**Figure 3: SurgencyP Theoretical and Empirical ICC**

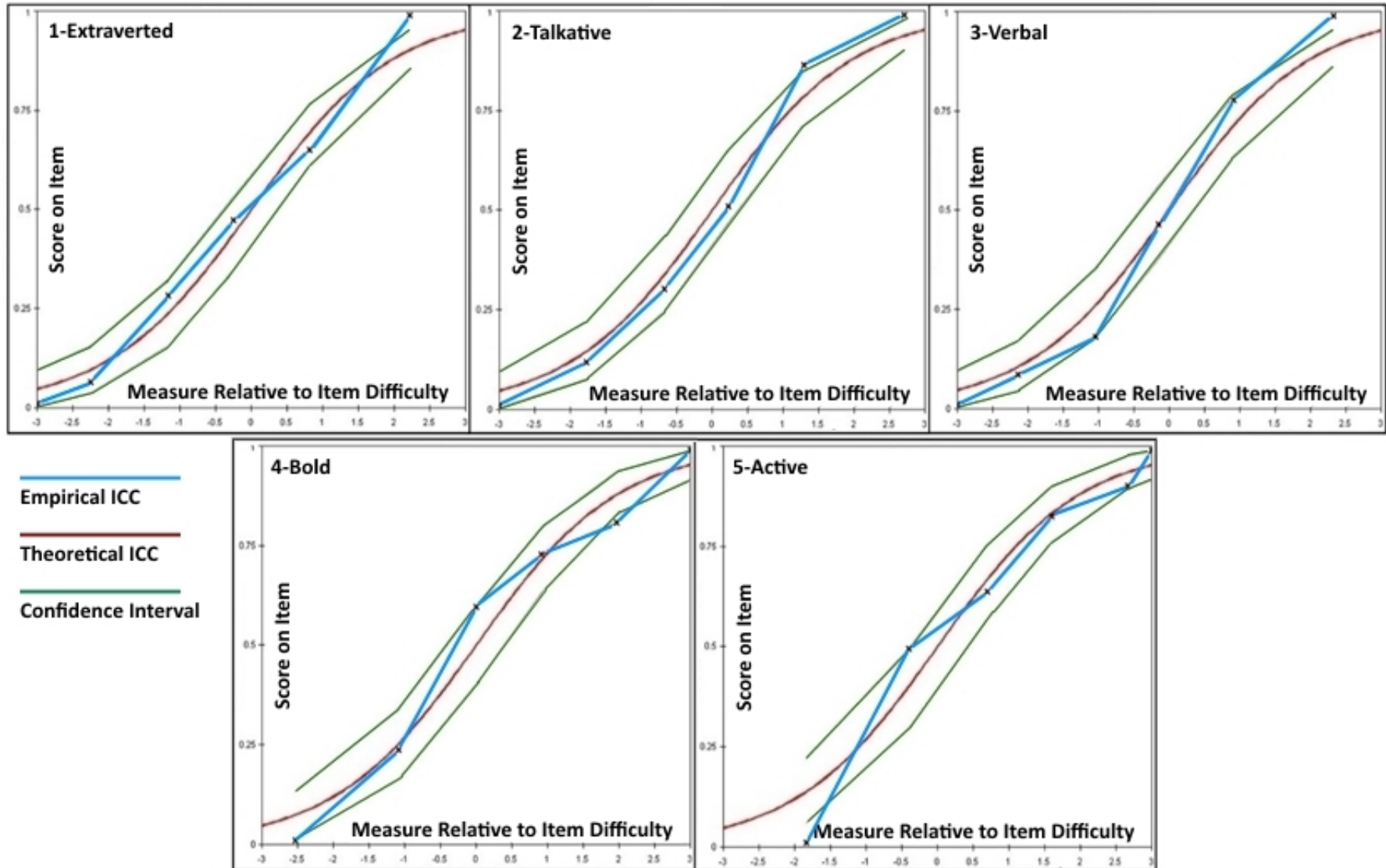
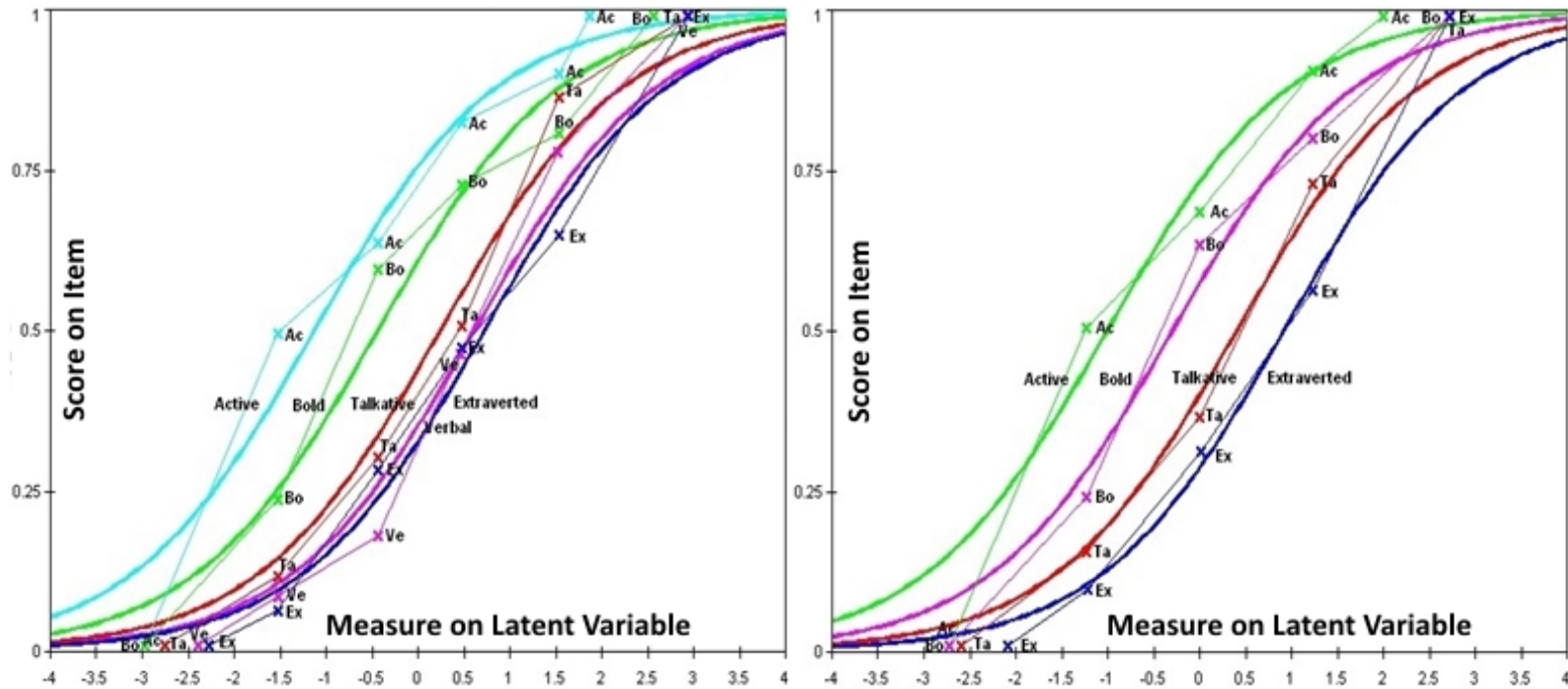


Figure 4: SurgencyP Initial & Reduced Multiple ICC





*SurgencyP*'s empirical ICCs approximate reasonably well their theoretical ideals. Jagged curves are quite normal (Linacre 2009b). For all traits, allocation probability is directly related to respondent propensity and trait intensity. A direct relationship exists with respect to the latent variable. Empirical ICCs also tend to be within their 95% confidence intervals. Exceptions are extreme respondents, those assigning all or no traits. This further confirms model compliance. However, not all *SurgencyP* traits perform equally well. For example, *active* has its response categories concentrated in the upper allocation probability range. This means that *active* is relatively frequently allocated by respondents.

### 2.6 Concurrent Item Functioning

Multiple ICCs can be simultaneously displayed. This reveals how items operate together, as a scale. The left side of Figure 4 shows theoretical and empirical ICCs for *SurgencyP*'s five items. As before, the vertical axis indicates allocation probability. The horizontal axis indicates the difference between respondent propensity and item intensity. The thick solid lines are theoretical ICCs. The thin jagged ones are empirical ICCs. *SurgencyP*'s five ICCs are intensity ordered along the horizontal axis. *Active*, the most attributed item, and thus the least intense one, is to the far left. *Bold*, *talkative* and *verbal* of intermediate intensity follow. *Extraverted*, the least attributed item, and thus the most intense one, is to the far right.

ICCs should be well and evenly spaced. Each should cover a specific intensity range along the measurement continuum (Wright 1999). *SurgencyP* ICCs are somewhat evenly spaced. The only exception is *verbal*, wedged between *talkative* and *extraverted*. Being almost on top of *extraverted* makes *verbal* problematic from a position perspective. From a conceptual perspective, *verbal* is close to *talkative* and should indeed be adjacent to it. However, based on the initial construct map, *verbal* is less intense than *talkative* and should thus be to the latter's left, not its right. People are first *verbal*, then *talkative*. The inverse ordering is likely due to respondents being more familiar with *talkative* than with *verbal*. Being more familiar, *talkative* is assigned more often making it less intense than it should be, hence its anomalous position. When ICC are close together, some items are likely redundant, able to be culled without sacrificing overall scale integrity (Bond and Fox 2015). Since from a position and content perspective *verbal* is problematic, it was deleted. This improved *SurgencyP*'s inter-curve spacing providing a clearer measure, as shown in Figure 4 on the right.

### 3. Refined Measure Statistics

Culling *verbal* had other positive effects: It concentrated *SurgencyP* statistics further around ideal values: Maximum *outfit* MNSQ and ZSTD decreased from borderline values of 1.19/1.9 to the more acceptable 1.03/0.4. Maximum *infit* values decreased from 1.11/1.8 to 1.05/0.8. Minimum *outfit* MNSQ and ZSTD increased from 0.88/-2.1 to 0.95/-0.9. Minimum *infit* values increased from 0.89/-2.4 to 0.95/-1.1. Improved fit was confirmed by outfit and infit MNSQ standard deviations decreasing from 0.13/0.09 to 0.03/0.03. Also by PMC range tightening and shifting upwards from 0.63/0.72 to 0.67/0.71.

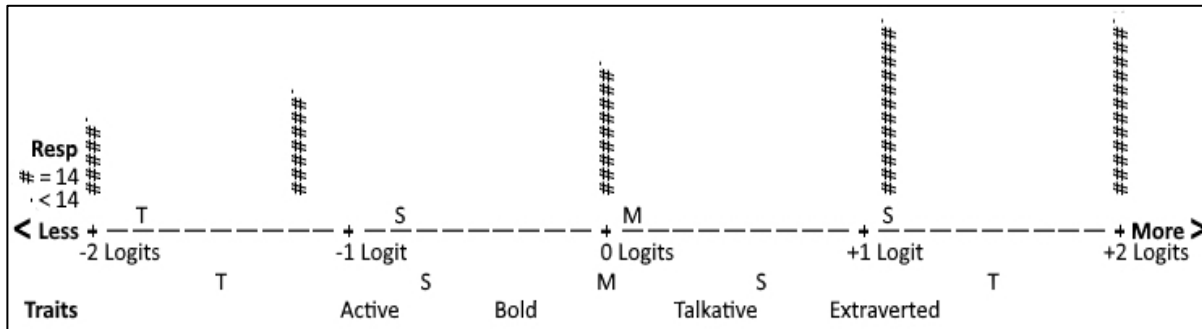
Measure *precision* is indicated by model error (Linacre 2009b). *SurgencyP*'s average error was 0.11 with a standard deviation of 0.0, indicating a rather precise measure (Linacre 2009c). Measure *reliability*, how consistent item hierarchies are over different applications, ranges from 0 to 1 (Bond and Fox 2015). *SurgencyP*' reliability was 0.97, very good. However, reliability may be overstated when error is small, as in this case. The *separation* coefficient resolves this by

confirming measure reliability. Separation indicates true variance opposed to error variance, ranging from 1.0 (50% reliability) to 9.0 (99% reliability) (Linacre 2009c). *SurgencyP*'s real and modelled separations of 6.17 and 6.23 support, though moderate, the former high reliabilities. In sum, *Surgency P* is a reasonably robust BP scale.

4. Item Person Map

The *SurgencyP* scale can be graphically represented. This is done by transforming the vertical ICC axis from probabilities into logarithmic units (logits). ICC curves thus become parallel lines, each covering a specific point along the construct's intensity continuum (Wright 1999). Figure 5, below, shows *SurgencyP*'s item-person map. It is akin to having a ruler of sorts through which BP item intensity and respondent propensity may be measured. Down the map's middle are scale intensities expressed in logarithmic units (logits). Being an interval scale, distances between units are equal, able to be concatenated (Bond and Fox 2015). Above the central axis are respondents. These are grouped into trait attribution categories. The five respondent categories, opposed to a continuous distribution, follow having four items. Respondents more prone to allocate *SurgencyP* traits are located towards the right, those less inclined to the left. Below the central axis are the four *SurgencyP* traits. More intense traits, allocated less frequently, are towards the right. Less intense traits, allocated more frequently, to the left. Average *SurgencyP* trait intensity (M) serves as the scale's zero point. Spread is indicated by one (S) and two (T) standard deviations.

Figure 5: SurgencyP Item-Person Map



5. Rasch Brand Personality Scales

Following the same process and criteria as with *SurgencyP*, nine additional Rasch BP scales were developed. Table 5, below, summarizes the scales' final items with their respective logit intensities.

The Rasch BP Scales (R-BPS) developed overcome key limitations of conventional instruments. First is the issue of validity. As previously noted, most BP instruments developed to date do not refer to mental characteristics in a strict sense. Broad construct definitions have led to the incorporation of physical characteristics, demographics, and socio-cultural evaluations, all outside human personality's realm, as understood by psychology (Allport 1938; Azoulay and Kapferer 2003). In contrast, the R-BPS derive from a BP definition and operationalization consistent with the biophysical notion of personality. They refer to personality, nothing else.

**Table 5: Rasch Brand Personality Scales**

<b>Surgency</b>	<b>Agreeableness</b>	<b>Conscientiousness</b>	<b>Emotional Stab.</b>	<b>Intellect</b>
<b>Positive</b>	<b>Positive</b>	<b>Positive</b>	<b>Positive</b>	<b>Positive</b>
Extraverted (0.91)	Generous (0.97)	Steady (1.18)	Unexcitable (0.46)	Deep (1.72)
Talkative (0.40)	Kind (0.67)	Efficient (0.00)	Unenvious (0.25)	Artistic (0.61)
Bold (-0.30)	Considerate (0.42)	Organized (-1.19)	Undemanding (0.23)	Bright (-0.96)
Active (-1.01)	Trustful (-0.66)		Relaxed (-0.94)	Intellectual (-1.37)
	Helpful (-1.39)	<b>Negative</b>		
<b>Negative</b>		Disorganized (0.55)	<b>Negative</b>	<b>Negative</b>
Shy (1.07)	<b>Negative</b>	Sloppy (0.18)	Fearful (0.49)	Uninquisitive(0.51)
Timid (0.49)	Rude (0.67)	Undependable (-0.74)	Nervous (0.39)	Unreflective (0.10)
Withdrawn (-0.01)	Distrustful (0.15)		Anxious (0.10)	Uncreative (0.04)
Introverted (-0.38)	Harsh (0.08)		Insecure (0.01)	Unsophisticated (-0.03)
Reserved (-1.18)	Uncooperative (-0.09)		Jealous (-1.0)	Shallow (-0.62)
	Cold (-0.81)			

**Note:** The average correlation between positive and negative domain sub-scales remained similar, -0.385 before refinement, -0.377 thereafter. Noteworthy is EM's P/N correlation improvement from 0.037 to -0.037. Other scale P/N correlations diminished slightly, Surg. from -0.328 to -0.320; Agree. from -0.581 to -0.531; Cons. from -0.570 to -0.512; and Intel. from -0.483 to -0.486. These figures support subscale independence.

Second is the issue of actual measurement. Most BP instruments are taxonomies, not scales. They lack intrinsic intensities and units, unable to properly quantify personality variables. The R-BPS resolve this by approximating the fundamental requirements of proper measures: *Unidimensionality*, referring to single, albeit general constructs; *invariance*, being rather consistent regardless of respondents and stimuli; and *concatenation*, units able to be added with one another. Unlike previous instruments, the R-BPS measure BP in a true sense. Brands' attributed personality may now be properly quantified.

Third is the issue of comprehensiveness. Most BP instruments consist of positive traits only. This also stems from Aaker J (1997), who argues that BP instruments should focus on assessing the extent to which consumers approach, rather than avoid, brands. However, human personality consists of both positive and negative traits (Cervone and Pervin 2008). Including positive BP traits only leads to incomplete assessments (Geuens, Weijters, and De Wulf 2009). If BP's complex nature is to be truly understood, instruments contemplating both positive and negative traits must be developed (Azoulay 2007; Stapley 1996). The R-BPS also overcomes this.

Fourth is the issue of taxonomical congruence with human personality. The Big Five are personality's fundamental dimensions, irrespective of culture, gender and age group (Cervone and Pervin 2008). However, most BP instruments developed to date only marginally resemble the Big Five. They mix items from multiple dimensions, with some BP dimensions being completely beyond the scope of human personality. Based on human personality, BP's structure should mirror the Big Five (Geuens et al. 2009). The R-BPS achieves this.

Lastly is the issue of predictive power. Most BP instruments are descriptive and refer only to respondents' past performance. However, measures should not only be descriptive, but also predictive (Weiss and Yoes 1991), especially personality measures, whose main purpose is to forecast outcomes (Chaplin, John, and Goldberg 1988). The R-BPS are probabilistic, and they predict performance at both the respondent and item level.

## Limitations

Despite overcoming conventional BP instrument weaknesses, the R-BPS do have their own limitations. Mental characteristics are continua (Eysenck 1965; Thurstone 1934) and items covering these continua should be conceptually cumulative, progressively reflecting more of the variable in question (Smith AB et al. 2006). However, the R-BPS do not always have progressively-ordered items. Conceptually more intense items are sometimes below less intense ones.

Items should also cover the entire response spectrum, be evenly spaced, and align with respondents. Measurement error otherwise increases as uncovered or misaligned respondents need estimation (Bond and Fox 2015). The R-BPS cover about half their response spectra, and are only suited for moderate response situations. Item spacing also shows some clusters and gaps, with items not always aligned with respondents.

## Future Research

The above limitations (non-progressively ordered items, spectrum coverage, spacing, and alignment) might be attributed to scales comprising few final items. All else being equal, additional items tend to improve scale properties: They extend range and normalize spacing, alignment and order (Linacre 2009; Smith AB et al. 2006). However, more fundamental factors were also considered.

*Item Source:* Operationalisation in this study used Goldberg's 100 Markers. However, these were factor-analytically derived. Item quality was thus limited by FA's reliability maximization criteria which narrows construct scope (Duncan 1984). To reduce this problem, larger taxonomies might instead be used, or better yet, personality trait lists. While comprehensive, extensive item sets are also impractical. Instead of developing scales for all BP dimensions at once, as presently done, future efforts might instead concentrate on single dimensions. This would take full advantage of a list's richness, while keeping trait numbers manageable.

*Scale Dimensionality:* Multiple constructs within a single measure distort results (Bond and Fox 2015). The R-BPS in general refer to individual constructs, but they are not entirely unidimensional. They likely contain multiple sub-dimensions, each pertaining to different, more-specific constructs. That R-BPS items do not order progressively is likely caused by combined latent sub-dimensions (Linacre 2009b). Future Rasch scaling efforts should thus first uncover variables' complete taxonomies and only then develop corresponding scales. Addressing dimensionality requires a substantial number of initial items, large enough to afford multiple dimensional splits, cull non-functioning items, and then still have a broad and densely populated scale. This further supports using large initial trait sets focused on single dimensions.

*Item clarity:* Goldberg's Markers were selected due to being relatively understandable. However, they still included adjectives that were somewhat obscure (for example, *imperturbable*), subject to interpretation (for example *bold*, meaning *confident* or *imprudent*), or whose subtle differences might not be clear to respondents (for example, *verbal* versus *talkative*). This likely impacted item progression. To further improve future efforts, more easily-understood items should be used. Culling obscure and confusing traits might be done by cross referencing against other sources or

through preliminary research. Respondent understanding can also be improved by providing a context, be it formulating items as sentences, using bipolar pairs, or providing adjective definitions.

*Stimuli:* Not only brands but entire product categories have characteristic personalities (Alt and Griggs 1988; Venable *et al.* 2005). For example, business schools are typically perceived as *competent* (Opoku, Abratt, and Pitt 2006), tourism destinations as *exciting* (Murphy *et al.* 2007), and shampoos as *gentle* (Smit, Bronner, and Tolboom 2007). Having used only technology brands likely emphasized particular characteristics, impacting item progression: Technology is associated with performance and innovation. This might explain why with *SurgencyP* *active* and *bold* were more frequently attributed, hence less intense, than *talkative* and *extraverted*, arguably characteristics of lesser importance for technology. To counter this, future Rasch scaling efforts should include stimuli from a range of product categories.

*Instrument:* To reduce item skipping and improve data quantity/quality, checklist answer options were extended from single “yes” tick boxes to double “yes” or “no” ones. Instructions also requested that respondents answer all items. These measures appear to have had an undesirable side effect: Traits, whose meaning was not clear nor applicable to the brands, were forced into the model affecting results. Future Rasch efforts should thus extend checklist answer options to include a “don’t know/not applicable” tick box. This will slightly increase respondent effort, but will help exclude incorrect answers, thus improving data quality.

## Conclusion

This research set out to explore Rasch-based scaling towards the development of BP scales that more closely approximate MT guidelines. This was accomplished. The R-BPS are a solid first approximation to proper measures within the field of BP. A series of implementation-related issues were identified and discussed. Suggestions as to how future efforts might be improved were also provided, setting the stage for subsequent refinement.

Empirical results should support extant theory. But it is also valuable for them to conflict as this elucidates important issues. The only way fields can truly develop is through the resolution of discrepancies (Linacre 2009b; Wright 1999). From this perspective, and despite the scale’s limitations, the present study makes useful contributions. It is a first step in a new line of enquiry, providing foundations for further research.

Studies that extend fields’ methods into new domains are valuable (Busenitz *et al.* 2014). Conventional approaches, for example, Churchill (1979), assume that constructs only have horizontal content breadth. Present results suggest otherwise. In line with what psychology has long hinted, (e.g. Eysenck 1965), constructs might also operate along vertical intensity continua encompassing content strength. Researchers are thus encouraged to expand their view of constructs. Approaching them from an intensity perspective, not only breadth, is likely fertile ground for future research.



## References

- Aaker D, Kumar V, Day GS, & Lawley M (2005). *Marketing Research - Pacific Rim Edition*. Milton, Qld: John Wiley & Sons, Australia.
- Aaker J (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, 34(3), 347-356.
- Allport GW (1938). *Personality - A Psychological Interpretation*. London, UK: Constable & Co.
- Alt M, & Griggs S (1988). Can a brand be cheeky? *Marketing Intelligence and Planning*, 6(4), 9-16.
- Andrich D (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38(3), 665-680.
- Andrich D (1985). An Elaboration of Guttman Scaling with Rasch Models for Measurement. *Sociological Methodology*, 15(1), 33-80.
- Andrich D (1988). *Rasch Models for Measurement* (First ed.). Newbury Park, CA: Sage Publications.
- Austin JR, Siguaw JA, & Mattila A (2003). A re-examination of the generalizability of the Aaker brand personality measurement framework. *Journal of Strategic Marketing*, 11(2), 77-92.
- Azoulay A (2007). Does Brand Personality Make a Difference? (pp. 52). Paris: HEC School of Management.
- Azoulay A, & Kapferer J-N (2003). Do Brand Personality Scales Really Measure Brand Personality? . *Journal of Brand Management*, 11(2), 143-155.
- Belk RW, & Tumbat G (2005). The Cult of Macintosh. *Consumption, Markets and Culture*, 8(3), 205-217.
- Bond TG, & Fox CM (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd. ed.). New York, NY: Routledge.
- Bosnjak M, Bochmann V, & Hufschmidt T (2007). Dimensions of Brand Personality Attributions: A Person-Centric Approach in the German Cultural Context. *Social Behavior and Personality*, 35(3), 303-316.
- Brody N (1994). Traits. In VS Ramachandran (Ed.), *Encyclopaedia of Human Behaviour* (Vol. 4, pp. 419-425). San Diego: Academic Press.
- Busenitz LW, Plummer LA, Klotz AC, Shahzad A, & Rhoads K (2014). Entrepreneurship Research (1985–2009) and the Emergence of Opportunities. *Entrepreneurship Theory and Practice*, 38(5), 981-1000.
- Cattell RB (1946). *Description and Measurement of Personality*. Yonkers, New York: World Book Company.
- Cervone D, & Pervin LA (2008). *Personality - Theory and Research* (10th ed.). Hoboken, NJ: John Wiley & Sons.
- Chaplin WE, John OP, & Goldberg LR (1988). Conceptions of States and Traits: Dimensional Attributes With Ideals as Prototypes. *Journal of Personality and Social Psychology*, 54(4), 541-557.
- Churchill GA (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16(February), 64-73.
- Conejo F, & Wooliscroft B (2015). Brands Defined as Semiotic Marketing Systems. *Journal of Macromarketing*, 35(3), 287-301.
- Cronbach LJ (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3), 297-334.

- De Battisti F, Nicolini G, & Salini S (2005). The Rasch Model to Measure Service Quality. *The ICFAI Journal of Services Marketing*, 3(3), 58-80.
- De Chernatony L, & McWilliam G (1990). Appreciating Brands as Assets Through Using a Two Dimensional Model. *International Journal of Advertising*, 9(2), 111-119.
- Deater-Deckard K, Mullineaux PY, Petrill SA, & Thompson LA (2009). Effortful Control, Surgency, and reading skills in middle childhood. *Reading and Writing*, 22(1), 107-116.
- DeVellis RF (2012). *Scale Development – Theory and Applications* (3rd Ed.). Thousand Oaks, CA: Sage Publications.
- Duncan OD (1984). *Notes on social measurement: Historical and critical*. New York: Russell Sage Foundation.
- Ewing M, Salzberger T, & Sinkovics R (2005). An Alternate Approach to Assessing Cross Cultural Measurement Equivalence in Advertising Research. *Journal of Advertising*, 34(1), 17-36.
- Eysenck HJ (1965). *The Structure of Human Personality*. London: Methuen & Co.
- Gaito J (1980). Measurement Scales and Statistics: Resurgence of an Old Misconception. *Psychological Bulletin*, 87(3), 564-567.
- Ganglmair-Wooliscroft A (2010). *Ownership Patterns for Durable Goods: A Preliminary Investigation of two New Zealand Lifestyle Segments*. Paper presented at the Exploring the Frontiers of Macromarketing: 35th Annual Macromarketing Conference, Laramie, WY.
- Ganglmair-Wooliscroft A, & Lawson R (2003). Advantages of Rasch Modelling for the Development of a Scale to Measure Affective Response to Consumption. *European Advances in Consumer Research*, 6, 1-6.
- Ganglmair-Wooliscroft A, & Wooliscroft B (2010). Different Englishes? Investigating Equivalency of the Affective-Response-to-Consumption Scale Amongst Geographically Disparate Groups of English Speakers. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 23(137-155).
- Ganglmair-Wooliscroft A, & Wooliscroft B (2013). A Cross-Cultural Application of the Affective Response to Consumption Scale: Investigating US-American and Austrian Passengers on Long-Haul Flights. *Journal of Business Research*, 66(6), 765-770.
- Geuens M, Weijters B, & De Wulf K (2009). A New Measure of Brand Personality. *International Journal of Research in Marketing*, 26, 97-107.
- Goldberg LR (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1), 26-42.
- Goldberg LR, & Digman JM (1994). Revealing Structure in the Data: Principles of Exploratory Factor Analysis. In S Strack & M Lorr (Eds.), *Differentiating Normal and Abnormal Personality* (pp. 216-242). New York: Springer.
- Hambleton RK, & Swaminathan H (1985). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer-Nijhoff Publishing.
- Jang KL, Livesley WJ, & Vernon PA (1996). Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study. *Journal of Personality*, 64(3), 577-592.
- Kaiser HF (1970). A Second Generation Little Jiffy. *Psychometrika*, 35(4), 401-415.
- Kuhn TS (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Ladik DM, & Stewart DW (2008). The Contribution Continuum. *Journal of the Academy of Marketing Science*, 36, 157-165.
- Laurent G, Kapferer J-N, & Roussel F (1995). The Underlying Structure of Brand Awareness Scores. *Marketing Science*, 14(3), 170-179.

- Linacre JM (1992). *Many facet Rasch measurement*. Chicago: MESA Press.
- Linacre JM (2003). Size vs. Significance: Infit and Outfit Mean-Square and Standardized Chi-Square Fit Statistic. *Rasch Measurement Transactions*, 17(1), 918, accessed online at [www.rasch.org/rmt/rmt171n.htm](http://www.rasch.org/rmt/rmt171n.htm).
- Linacre JM. (2009). WINSTEPS Rasch Measurement Computer Program (Version 3.69.0). Chicago: Winsteps.com.
- Linacre JM (2009b). Practical Rasch Measurement Core Topics - Online Course.
- Linacre JM (2009c, June 27-July 27). [Personal Communication - Practical Rasch Measurement Core Topics Discussion Board and Emails].
- Linacre JM. (2014). WINSTEPS Rasch Measurement Computer Program (Version 3.81.0). Chicago: Winsteps.com.
- Low MB, & MacMillan IC (1988). Entrepreneurship: Past research and future challenges. *Journal of Management*, 14(2), 139-161.
- Masters G (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Meads DM, & Bentall RP (2008). Rasch analysis and item reduction of the hypomanic personality scale. *Personality and Individual Differences*, 44(8/June), 1772-1783.
- Michell J (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355-383.
- Murphy L, Moscardo G, & Benckendorff P (2007). Using Brand Personality to Differentiate Regional Tourism Destinations. *Journal of Travel Research*, 46(5), 5-14.
- Nijsten T, Unaeze J, & Stern RS (2006). Refinement and reduction of the impact of psoriasis questionnaire: Classical test theory vs. Rasch analysis. *British Journal of Dermatology*, 154(4), 692-700.
- OED. (2016). Various Definitions. Retrieved November 5th, from <https://en.oxforddictionaries.com/>
- Opoku R, Abratt R, & Pitt L (2006). Communicating brand personality: Are the websites doing the talking for the top South African business schools? *Journal of Brand Management*, 14(1), 20-39.
- Peabody D, & Goldberg LR (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, 57, 552-657.
- Rasch G (1960/1993). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: MESA Press.
- Romaniuk J (2008). Comparing Methods of Measuring Brand Personality Traits. *Journal of Marketing Theory and Practice*, 16(2), 153-161.
- Salzberger T (1999). *How the Rasch model may shift our perspective of measurement in marketing research*. Paper presented at the Australia and New Zealand Marketing Academy Conference (ANZMAC), Sydney.
- Salzberger T (2000). *An alternative way of establishing measurement in marketing research—its implications for scale development and validity*. Paper presented at the Australia and New Zealand Marketing Academy Conference (ANZMAC), Gold Coast, Australia.
- Salzberger T (2009). *Measurement in Marketing Research: An Alternative Framework*. Cheltenham, UK: Edward Elgar.
- Salzberger T, & Koller M (2013). Towards a new paradigm of measurement in marketing. *Journal of Business Research*, 66(9), 1307-1317.

- Salzberger T, Newton FJ, & Ewing MT (2014). Detecting gender item bias and differential manifest response behavior: A Rasch-based solution. *Journal of Business Research*, 67(4), 598–607.
- Saucier G (1994). Mini Markers: A Brief Version of Goldberg's Unipolar Big Five Markers. *Journal of Personality Assessment*, 63(3), 506-516.
- Schumacker RE, & Linacre JM (1996). Factor analysis and Rasch analysis. *Rasch Measurement Transactions*, 9(4), 470, accessed online at [www.rasch.org/rmt/rmt494k.htm](http://www.rasch.org/rmt/rmt494k.htm).
- Shimp TA, & Sharma S (1987). Consumer ethnocentrism: construction and validation of the CETSCALE. *Journal of Marketing Research*, 24(August), 280-289.
- Smit EG, Bronner F, & Tolboom M (2007). Brand relationship quality and its value for personal contact. *Journal of Business Research*, 60, 627–633.
- Smit EG, Van Den Berge E, & Franzen G (2003). Brands are just like real people! The development of SWOCC's Brand Personality Scale. In F Hansen & LB Christensen (Eds.), *Branding and Advertising* (pp. 22-43). Copenhagen: Copenhagen Business School Press.
- Smith AB, Wright EP, Rush R, Stark DP, Velikova G, & Selby PJ (2006). Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. *Psycho-Oncology*, 15(9), 817-827.
- Smith AM (1999). Some Problems When Adopting Churchill's Paradigm for the Development of Service Quality Measurement Scales. *Journal of Business Research*, 46(2), 109–120.
- Smith RM (1996). A Comparison of Methods for Determining Dimensionality in Rasch Measurement. *Structural Equation Modeling*, 3(1), 25-40.
- Smith RM, Schumacker RE, & Busch MJ (1995). *Using Item Mean Squares to Evaluate Fit to the Rasch Model*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Soutar GN, Bell RC, & Wallis YM (1990). Consumer Acquisition Patterns for Durable Goods: A Rasch Analysis. *Asia Pacific International Journal of Marketing*, 2(1), 31-39.
- Stapley LF (1996). *The Personality of the Organisation*. London: Free Association Books.
- Steinberg L, & Thissen D (1996). Uses of Item Response Theory and the Testlet Concept in the Measurement of Psychopathology. *Psychological Methods*, 1(1), 81–97.
- Stevens SS (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 677-680.
- Thurstone LL (1934). The Vectors of Mind. *Psychological Review*, 41, 1-32.
- Thurstone LL (1974). *The Measurement of Values*. Chicago: University of Chicago Press.
- Tull DS, & Hawkins DI (1990). *Marketing Research - Measurement and Method* (5th ed.). New York: Macmillan Publishing Company.
- Venable BT, Rose GM, Bush VD, & Gilbert FW (2005). The Role of Brand Personality in Charitable Giving: An Assessment and Validation. *Academy of Marketing Science Journal*, 33(3), 295-312.
- Wagh RF, & Chapman ES (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, 6(1), 80–99.
- Weiers RM (1988). *Marketing Research* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall International.
- Weiss DJ, & Yoes ME (1991). Item response theory. In RK Hambleton & JN Zaal (Eds.), *Advances in educational and psychological testing: theory and applications* (pp. 69-95). Boston: Kluwer Academic Publishers.

- Wooliscroft B, Ganglmair-Wooliscroft A, & Noone A (2014). The Hierarchy of Ethical Consumption Behavior: The Case of New Zealand. *Journal of Macromarketing*, 34(1), 57-72.
- Wright BD (1988). Georg Rasch and Measurement. *Rasch Measurement Transactions*, 2(3), 25-32.
- Wright BD (1994). Comparing factor analysis and Rasch measurement. *Rasch Measurement Transactions*, 8(1), 350, accessed online at [www.rasch.org/rmt/rmt381r.htm](http://www.rasch.org/rmt/rmt381r.htm).
- Wright BD (1999). Fundamental measurement for psychology. In SE Embretson & SL Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know*. Hillsdale, NJ: Lawrence Erlbaum Associates - Accessed online as MESA Research Memorandum No.64 at [www.rasch.org/memo64.htm](http://www.rasch.org/memo64.htm).
- Wright BD, & Linacre JM (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Yamaguchi J (1997). Positive vs. Negative Wording. *Rasch Measurement Transactions*, 11(2), 567, accessed online at [www.rasch.org/rmt/rmt112h.htm](http://www.rasch.org/rmt/rmt112h.htm).
- Zwick R, Thayer DT, & Lewis C (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. *Journal of Educational Measurement*, 36(1), 1-28.