Behavior Research Methods
2010, 42 (4), 944-956
doi:10.3758/BRM.42.4.944

The many-facet Rasch model in the analysis of the go/no-go association task

Michelangelo Vianello and Egidio Robusto
University of Padua, Padua, Italy

This article provides a many-facet Rasch measurement (MFRM) analysis of go/no-go association task (GNAT)-based measures of implicit attitudes toward sweet and salty food. We describe the statistical model and the strategy we adopted to score the GNAT, and we emphasize that, when analyzing implicit measures, MFRM indexes have to be interpreted in a peculiar way. In comparison with traditional scoring algorithms, an MFRM analysis of implicit measures provides some additional information and suffers from fewer limitations and assumptions. MFRM might help to overcome some limitations of current implicit measures, since it directly addresses some known issues and potential confounds, such as those related to a rational zero point, to the arbitrariness of the metric, and to participants' task-set switching ability.

During the last 20 years, research on implicit methods has increased exponentially. Major implicit (or indirect) techniques include evaluative priming (EP; Fazio, Sanbonatsu, Powell, & Kardes, 1986) and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which has been followed in rapid succession by many other tests, such as the go/no-go association task (GNAT; Nosek & Banaji, 2001), the extrinsic affective Simon task (EAST; De Houwer, 2003), the affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005), the single-category–IAT (SC–IAT; Karpinski & Steinmen, 2006), and the sorting paired features (SPF; Bar-Anan, Nosek, & Vianello, 2009). Although these techniques may imply very different procedures from one another, they all share the aim of circumventing the influence of corrective processes involved in explicit measures (e.g., questionnaires), which may be heavily influenced by social desirability or impression management strategies. Indeed, implicit techniques do not rely on introspection. On the contrary, they provide behavioral measures of association strengths among mental representations, and they all rely on the assumption that the processing of a stimulus increases the accessibility of associated concepts (Higgins, 1996).

For example, in an IAT for measuring implicit racial bias (one of the most common applications), participants categorize words into superordinate categories in two different sorting conditions. In one condition, participants categorize items representing *whites* (e.g., faces of white people) and *good words* (e.g., *good*, *beautiful*) with one response key, while categorizing items representing *blacks* (faces of black people) and *bad words* (*bad*, *evil*) using another response key. In the other condition, participants categorize the same stimuli, but in different pairs: *white* and *bad* items are categorized with one key, whereas *black* and *good* items are categorized with the other. The first condition (*white–good*) is typically easier than the second (*white–bad*; see Nosek, Greenwald, & Banaji, 2006). The individual difference in speed and/or accuracy between conditions is interpreted as a measure of participants' implicit preference for whites over blacks, which has often been interpreted as a measure of implicit prejudice. Although existent measures of association strengths use distinct procedures and may tap different associative processes, they all derive their evaluations from comparisons between participants' performances on different categorization or recognition tasks. For instance, individual scores on a race–EP are obtained by comparing responses to targets that were preceded by a stimulus priming the concept *black* with responses that were preceded by a neutral prime or by a *white* prime. In the logic of response competition tasks, rather than of sequential priming, the GNAT derives individual scores according to signal detection theory (Green & Swets, 1966). Hence, the individual measure of implicit association—which is called sensitivity (d')—is computed by subtracting the standardized proportion of hits (correct responses to targets) from the standardized proportion of false alarms (incorrect responses to distractors). The d' represents the individual's ability to discriminate signals (target stimuli) with noise from noise alone (distractor stimuli).

All implicit techniques are characterized by specific strengths and limitations. GNAT, EP, SC–IAT, AMP, and SPF have an advantage over the IAT in that they provide measurements of implicit associations that are not relative. For example, an IAT on racial prejudice provides a relative measure of participants' implicit association of white people and good, relative to the association between black people and bad. Nonetheless, GNAT and EP are character-

M. Vianello, michelangelo.vianello@unipd.it

ized by a notable lack of reliability, as compared both with the IAT and with other explicit techniques. The GNAT has values of internal consistency between .1 and .3 (Nosek & Banaji, 2001), whereas EP has even lower values (Bosson, Swann, & Pennebaker, 2000; Fazio & Olson, 2003; Olson & Fazio, 2003). Furthermore, the d' analysis requires that hit and false alarm rates be neither 0% nor 100%, and corrections have to be applied in these cases (Banaji & Greenwald, 1995). In addition, this analysis cannot be applied to participants with an error rate higher than 50%. Lastly, d' values are differential scores, which have been often criticized because of their low reliability (difference scores suffer from a lack of reliability, which is a function of the correlation between the original variables; see Nunnally & Bernstein, 1994, for an analytical demonstration).

This study introduces an alternative model to analyze GNAT-based measures of implicit associations. The next section introduces the model.

The MFRM and Its Main Advantages

The many-facet Rasch model (Linacre, 1989) derives from the simple logistic model (SLM; Rasch, 1960/1980). Given that x_{ni} is a response to a test, which is 1 if the response is correct or 0 if the response is incorrect, β_n is the ability of the individual n, and δ_i is the difficulty of the item i, the SLM takes the following mathematical form:

$$P\left(X_{ni} = x_{ni} \middle| \beta_n, \delta_i\right) = \frac{\exp\left[x_{ni}\left(\beta_n - \delta_i\right)\right]}{1 + \exp\left(\beta_n - \delta_i\right)}. \qquad (1)$$

We can note that the model expresses, as for a logistic regression, the probability of obtaining a certain response as a function of the ability of the individual and of item difficulty ($\beta_n - \delta_i$). The more (or less) able the individual is and the easier (or more difficult) the item is, the more (or less) probable it will be that a correct response will be obtained.

If, by using Equation 1, we intend to calculate the probabilities associated with the events *correct response* and *incorrect response*, then we obtain, respectively,

$$P\left(X_{ni} = 1 \middle| \beta_n, \delta_i\right) = \frac{\exp\left(\beta_n - \delta_i\right)}{1 + \exp\left(\beta_n - \delta_i\right)} \qquad (2)$$

and

$$P\left(X_{ni} = 0 \middle| \beta_n, \delta_i\right) = \frac{1}{1 + \exp\left(\beta_n - \delta_i\right)}. \qquad (3)$$

By considering Equations 2 and 3 and calculating the logarithm, we obtain

$$\begin{aligned} \ln \frac{P\left(X_{ni} = 1 \middle| \beta_n, \delta_i\right)}{P\left(X_{ni} = 0 \middle| \beta_n, \delta_i\right)} &= \ln \frac{\exp\left(\beta_n - \delta_i\right) / \left[1 + \exp\left(\beta_n - \delta_i\right)\right]}{1 / \left[1 + \exp\left(\beta_n - \delta_i\right)\right]} \\ &= \beta_n - \delta_i. \qquad (4) \end{aligned}$$

As is evident in this formulation, we can introduce further parameters (facets) that all lie on the same trait. It therefore follows, as far as the application of the model to implicit techniques is concerned, that it is possible to introduce a third parameter that accounts for the probability of obtaining a given response on any given association task. This parameter is defined as the *condition of association* (γ_j), which assumes a different value for each different task (often called critical blocks) that is analyzed in the same model. For instance, in a two-block GNAT measuring implicit prejudice toward black people, one critical block (condition of association) would ask the participant to identify stimuli representing black people and good words ($j = 1$), and another one would ask the participant to identify black people and bad words ($j = 2$). In cases in which the GNAT employed four critical blocks, $j = 3$ would represent, for example, the block *white people* and *good words*, and $j = 4$ would represent, for example, *white people* and *bad words*. Hence, the model assumes the following three-facet formulation:

$$\ln \frac{P\left(X_{nij} = 1 \middle| \beta_n, \delta_i, \gamma_j\right)}{P\left(X_{nij} = 0 \middle| \beta_n, \delta_i, \gamma_j\right)} = \beta_n - \delta_i - \gamma_j. \qquad (5)$$

The Rasch model parameters are additive, fully satisfying one of the essential requisites of interval measures, and they are based on the transformation of scores into a *logit* scale, a logarithmic transformation of the probability of producing a particular response, given certain conditions (participants' ability, stimuli recognizability, and conditions' difficulty). In Equation 5, the *logit* can be seen as the dependent variable, whereas the various factors (e.g., participants, items, and conditions) act as independent variables that influence (or control) the response.

The MFRM is a member of the Rasch family of models; therefore, it is characterized by specific objectivity (or relational invariance), linearity, and measurement units (for a discussion of these properties see, e.g., Andrich, 1988).

Specific objectivity (SO) is one of the most interesting features of Rasch models for the analysis of implicit measures. SO postulates that within the same frame of reference, the comparison between objects (e.g., the difficulty of conditions of association) should be independent from other objects (e.g., subjects' ability). A detailed explanation and proof of SO is provided in the Appendix.

In Rasch models, the marginal sums of the data matrix are *sufficient statistics* to estimate the parameters. In the case of Equation 5—in a participant × conditions × item data matrix in which the response of subject n to item i in condition j is x_{nij}, and where x_{nij} has a value of 0 or 1 if a wrong or a correct answer, respectively, is given—the sufficient statistics for parameter estimates are $\Sigma_i \Sigma_j\ x_{nij}$ for participants, $\Sigma_n \Sigma_j\ x_{nij}$ for stimuli, and $\Sigma_n \Sigma_i\ x_{nij}$ for conditions. Hence, sufficient statistics can be interpreted as accuracy scores.

In summary, there are many advantages to using the MFRM to analyze association strengths. (1) All the facets lie on a common dimension of categorization accuracy. (2) As a consequence of SO, the measures obtained by the model are sample-, item-, and condition-free (hence, any parameter estimate can be compared with any other). (3) Specific goodness-of-fit statistics allow us to assess the fit of the data to the model, and they help us to interpret the results; such statistics allow an examination of the data that not only is general or comprehensive, such as that available from internal consistency statistics, but also evaluates the fit of each single item, participant, or association condition. (4) The experimental procedure can be limited to the simple registration of a correct or an incorrect response and can promote interval measures that express the latent trait. (5) The estimation of the parameters and the calculation of the associated measurement errors provide a simple and direct means of determining the significance of the differences between the experimental conditions, which will then represent individual- and group-level measures of implicit associations. In the following section, we will provide an MFRM analysis of a GNAT study, highlighting how this model helps us to answer important research questions regarding, for instance, the significance of group- and individual-level scores of implicit association, the quality of the stimuli utilized, potential confounding factors (e.g., task-set switching ability), sample size's appropriateness, and many others.

METHOD

Participants, Materials, and Procedure

The study sample consisted of 60 psychology undergraduates from the University of Padua that participated in the study for no reward. The experimental procedure (Inquisit software) provided a GNAT for the evaluation of participants' associations among sweet food, salty food, good words, and bad words (evaluative attributes).

The GNAT is a single-category association task, in which participants are asked to "catch" words or images belonging to two categories (attribute and target) by pressing the space bar (go) and to ignore (no go) all other stimuli (distractors). In one of the two critical blocks (conditions of association), participants were asked to press the space bar if the stimulus in the center of the screen belonged to one of two categories indicated under the stimulus (*sweets* and *good words*) and to do nothing if the stimulus did not belong to either of these two categories. Distractors were bad words and images of salty food. In a second critical block, participants were asked to press the space bar if the stimulus belonged to the categories *sweets* and *bad words*. Distractors in this block were good words and images of salty food. The GNAT effect is based on the individual difference in performance (accuracy) between these two critical blocks. Two more critical blocks were provided to evaluate the implicit attitude toward salty food. Errors were followed by a red cross. The stimuli appeared in black on a white background. Figure 1 provides all stimuli used in the procedure. In this figure, uppercase labels (e.g., BEAUTIFUL) refer to words, and lowercase labels (e.g., Salty1.jpg) refer to photographs of actual products. The fixation point lasted 400 msec; the response windows lasted from 500 to 650 msec for distractors and from 700 to 850 msec for target stimuli. Following a three-wave longitudinal design, participants completed a first GNAT in Time 1, a second GNAT after a week (Time 2), and a third GNAT a month after Time 1.

The MFRM and the Analysis of the GNAT

We adopted a four-facet Rasch model. The dependent variable *accuracy* is represented by the value 0 in the case of an error and by the value 1 in the case of a correct response. The parameters β, δ, γ, and τ represent the location on the latent trait of *participants*, *stimuli*, *conditions* of association, and *time* when the measures were taken, respectively. The four conditions asked participants to identify sweet food and good words, sweet food and bad words, salty food and good words, and salty food and bad words.

RESULTS

Data analysis was completed using Facets v. 3.66.1 (Linacre, 2009a). All fit indexes of the model were satisfactory. The data log-likelihood chi-square is $L^2_{(26217)} = 19{,}054.16$ ($p > .99$). L^2 is an index of global fit that provides a test of the divergence between observed and expected scores, with degrees of freedom (*df*) equal to the number of observations less the number of free parameters (Fisher, 1970). This statistic often shows significant misfit, especially when the *df* is high. Moreover, when automatic association data are analyzed, participants are expected to perform differently across the conditions (i.e., critical blocks), and this situation increases the global misfit represented by the log-likelihood chi-square. For this reason, *infit* and *outfit* indexes are more useful. Infit and outfit represent the relationship between observed and model-derived response probabilities, and they have a range that goes from zero to infinity. Statistics equal to or near 1 indicate perfect correspondence between observed and expected values; statistics above 1 indicate the presence of greater variance than that modeled (noise); and statistics below 1 indicate the existence of lower variance in the data than that predicted by the model (muting). Infit/outfit values higher than 2 signal the presence of serious distortions in the data; values between 1.5 and 2 indicate the presence of distortions that do not, however, bias the overall goodness of fit of the measurement system; values between 0.5 and 1.5 indicate a good fit of the data to the model; and values less than 0.5 signal the presence of distortions capable of artificially inflating reliability measures and internal consistency, without altogether biasing the overall goodness of fit of the measurement system (Linacre, 2009a; Linacre & Wright, 1994). The difference between the infit and the outfit values derives from the way in which such statistics are calculated. Both are based on the differences, calculated for all responses to all stimuli for each participant, between observed responses and model-derived response probabilities. These residuals ($R_{nij} = E_{nij} - X_{nij}$) are given by the difference between the model-derived expected scores (E_{nij}) and the observed scores (X_{nij}), and they can be standardized by dividing them by the square root of the variance of the expected scores ($\sqrt{V_{E_{nij}}}$). Details on how E_{nij} and $V_{E_{nij}}$ are computed can be found in Myford and Wolfe (2003). The outfit statistic is a mean of the squares of the standardized residuals (i.e., a mean of the variances), whereas the infit statistic is calculated by weighting each squared standardized residual by the variance of (1) each participant, if the statistic is in regard to an item or elements of other *facets*, or (2) each item, if the statistic is in regard to a participant. For example, the outfit of a participant is computed according to Equation 6:

$$\text{Outfit} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I} Z^2_{R_{nij}}}{JI}, \tag{6}$$

where $Z_{R_{nij}}$ is the standardized difference between model-derived expected scores and those observed in the data

Measure's Logits *(Accuracy)*	Stimuli's *Difficulty*	Time *Difficulty*	Participants' *Ability*	Blocks' *Difficulty*	Measure's Logits (Accuracy)
More					More
			56		
1			17		1
			30 58 9		
			29		
			27 39 53		
	“Sweet9.jpg”		46 47 48 60		
			2 35 41 51		
			40 42 50 55		
	“Sweet7.jpg”		15 18 33 36		
			23 4		
0			10 13 16 25 31		0
			12 21 3 45 54		
	“SUCCULENT” “Salty8.jpg”		32 57 6	Salty + BAD	
	“SUBLIME” “Salty3.jpg” “Salty6.jpg”		1 22 37 5 7		
	“Sweet3.jpg” “PLEASANT” “Salty15.jpg” “Salty5.jpg”	T1	28 43 49	Sweet + GOOD	
	“NICE” “Sweet1.jpg” “Sweet10.jpg” “Sweet8.jpg” “HEAVEN” “PARADISE” “UNPLEASANT” “Salty7.jpg”		19 8		
	“ATTRACTIVE” “Sweet11.jpg” “Sweet13.jpg” “Sweet5.jpg” “EXCELLENT” “NASTY” “Salty14.jpg” “Salty2.jpg”				
	“Sweet12.jpg” “Sweet16.jpg” “Salty1.jpg” “Salty11.jpg”		14 24 44		
	“TASTY” “GOOD” “BAD” “DISGUSTING” “Sweet15.jpg” “BETTER” “PLEASURE” “COOL” “DISGUST” “Salty10.jpg” “Salty13.jpg” “Salty16.jpg”		11 26 34 59		
	“Salty4.jpg”				
	“AWFUL” “Sweet14.jpg” “Sweet4.jpg” “STINK” "UGLY" “STENCH” “FUNNY” “Salty9.jpg”	T2 T3			
–1	“Sweet6.jpg” “MARVELLOUS” “WORST” “WORSE” “REVOLTING” “Salty12.jpg”		20 38 52	Salty + GOOD	–1
	“HELL” “HATE” “BEAUTIFUL” “VOMIT”			Sweet + BAD	
	“NAUSEA”				
	“Sweet2.jpg”				
Less					Less
Measure's Logits *(Accuracy)*	Stimuli's *Difficulty*		Participants' *Ability*	Blocks' *Difficulty*	Measure's Logits (Accuracy)

Figure 1. Elements of the four facets on the latent trait “accuracy.” $L^2_{(26217)} = 19{,}054.16, p = 1$.

Table 1
The Association Conditions Measurement Report, Providing Estimates of Implicit Associations at the Group Level, Single and Group Fit Indexes, and Indexes of Separation Between Conditions

				Infit		Outfit	
Condition	Obs. Score	Measure	Model *SE*	MnSq	*Z*Std	MnSq	*Z*Std
Salty+Bad	5,382	−0.23	.03	1.02	1.00	1.07	2.6
Sweet+Good	5,538	−0.38	.03	1.00	0.10	1.01	2.0
Salty+Good	5,947	−0.96	.04	0.98	−0.50	0.92	−2.0
Sweet+Bad	5,974	−1.07	.04	0.99	−0.40	0.97	−0.7
Mean	5,710.3	−0.66	.04	1.00	0.10	0.99	0.0
SD	256.4	0.36	.00	0.01	0.60	0.06	1.7

Notes—*RMSE* = .04 (population); Adj *SD* = .36; *G* = 9.28; *H* = 12.70; *R* = .99; fixed (all same) $\chi^2(3) = 353.2, p < .001$.

(residuals), J is the number of conditions in the analysis, and I is the number of stimuli. The infit for the same participant is weighted, in order to give less importance to extreme scores, and is computed according to Equation 7:

$$\text{Infit} = \frac{\sum_{j=1}^{J}\sum_{i=1}^{I} Z^2_{R_{nij}} V_{E_{nij}}}{\sum_{j=1}^{J}\sum_{i=1}^{I} V_{E_{nij}}}. \quad (7)$$

Consequently, whereas the outfit statistic places greater emphasis on the residuals associated with responses that are farther from the measure of a given element, the infit statistic gives greater emphasis to those responses that are nearest to the measure of a given element (Bond & Fox, 2001; Wright & Masters, 1982).

In our data, for the facet association conditions, the infit ranges from 0.95 (salty+good) to 1.02 (salty+bad). For the time facet, infit ranges from 0.98 to 1.02; for the participants facet, infit values range from 0.94 to 1.05, and for the items facet, infit values range from 0.95 to 1.08, largely within the acceptable range. According to Smith (1996, 2002), we can interpret this result as evidence of unidimensionality of the latent trait. Because previous studies have shown that positive and negative formulations of items sometimes load on different dimensions, we deemed it important to provide further evidence of unidimensionality. Hence, following Linacre (2009b), we estimated individual abilities (β_n) on the categorization task in two different models. The first one included only positive words, and the second one included only negative words (alternatively, the same test could also have been run for sweet and salty food, yet we think the latter would be less relevant). The uncorrected zero-order Pearson correlation between the two series of participants' abilities is positive and close to perfection ($r > .99$). Finally, a principal components analysis was run on standardized residuals. The unidimensionality is supported by this analysis as well, since the largest eigenvalue is equal to 2.54.

In the estimated model, measurement error (i.e., the *SE* of the estimates) is quite low, both for the conditions and for the items (between .16 and .03), indicating a good level of measure accuracy. Figure 1 represents how the various elements of the four facets lie on the *categorization accuracy* latent trait. The first column concerns the axis of the latent trait on which the various measures lie, and the values displayed are on the logit scale. The second column shows the difficulty of the stimuli. The third column shows the difficulty of each observation (time), the fourth provides participants' ability, and the fifth column shows the location of the four conditions of the associations analyzed.

For the conditions facet (see Table 1), a chi-square may be used to test whether all elements of the facets (four in this case) have the same logit value. The *fixed* (all same) chi-square tests the hypothesis that all the elements of the facet have the same logit in the population, in relation to the measurement error (*SE*). Hence, it helps to reject the hypothesis that there is no group-level implicit association between targets and attributes. An approximation to the theoretical distribution of chi-square can be obtained as indicated in Equation 8:

$$\chi^2 = \sum_j \frac{\gamma_j^2}{SE_j^2} - \frac{\left(\sum_j \frac{\gamma_j}{SE_j}\right)^2}{\sum_j \frac{1}{SE_j^2}}, \quad (8)$$

where the statistics are calculated for the facet conditions, with $j = 1, \ldots, L$ and $df = L - 1$, where L is the number of elements in the facet.

In this case the chi-square value is 353.10 (with 3 *df*s and $p < .001$); thus, at least one association is significantly different from the others. In Rasch measures, it is possible to compare different logits by dividing their difference by the square root of the sum of their error variance:

$$t = \frac{\gamma_1 - \gamma_2}{\sqrt{SE_1^2 + SE_2^2}}. \quad (9)$$

The standardization provides a value belonging to the Student's t distribution, with *df*s equal to the sum of "free" observations of the two elements. In this case, standardized estimates help to compare the four different associations under investigation. In our data, the salty+bad association is the most difficult, and specifically, it is significantly more difficult than the salty+good condition [$t(1) = 14.60, p < .001$]. The sweet+good condition is more difficult than the sweet+bad condition [$t(1) =$

13.80, $p < .001$]. The sweet+good and salty+bad conditions [$t(1) = 3.00$, $p = .20$] and the sweet+bad and salty+good conditions share the same difficulty [$t(1) = 2.60$, $p > .23$]. Hence, we observed a positive implicit attitude toward salty food and a negative implicit attitude toward sweet food of approximately the same intensity. In order to further analyze the implicit associations we measured, it is possible to consider three different indexes: the G, H, and R statistics. These are based on the same information, but they highlight different aspects of it. For example, when the elements of the facets are expected to be homogeneous, R is most useful, since when it is less than .5, it indicates that any differences of logit can be completely attributed to measurement error (Linacre, 2009a). Alternatively, if differentiation is expected (as is the case when the aim is to discriminate between the abilities of individuals), the G and H indexes are more useful.

The *separation ratio* (G) represents a measure of the difference between the scores obtained by the elements of the facet in relation to their precision (Linacre, 2009a; Myford & Wolfe, 2003). It is expressed as the relationship between the "true" standard deviation (i.e, the standard deviation of the estimates corrected for measurement error: adj $SD = SD - RMSE^2$) and the average of the standard error of the elements ($RMSE$). Therefore, $G =$ (adj SD)/$RMSE$ (see Linacre, 2009b, for computational details). The separation ratio (G) is extremely important in the analysis of the critical blocks (e.g., good vs. bad) utilized in the experimental procedures. If only two conditions are included in the analysis, their separation ratio is a measure of the mean automatic association effect among participants. The G of the facet conditions can be interpreted as a measure of the sensitivity of the instrument, and therefore, it is the first index to look at, in, for example, a study in which groups are strongly polarized and the expected value is obviously elevated. The separation of the participant facet is not as important as in traditional tests, where it represents a measure of the resulting discrimination. In classical intelligence and attainment tests, we would expect a high person separation value. In the case of implicit measures, it is different, because the measure of association is based on a comparison (bias/interaction analysis) between the performance in one condition (e.g., *bad*) and that in another condition (e.g., *good*). In implicit techniques, the general level of performance (speed and accuracy of response) is not of direct interest. We could, in theory, obtain an optimal measure of implicit association even without discriminating between participants in terms of their ability in completing the tasks. G for the participants' facet simply gives us an idea of how difficult the procedure is, and, all things being equal, it is preferable to obtain a measure that is just as difficult for all the participants; therefore, we expect *low indexes of separation between participants*. As far as the facet item is concerned, G provides useful information concerning the degree to which the stimuli represent the trait examined.

The second statistic on separation that we describe is the *separation index* (H), which is very similar to G, since it defines the number of different groups (heterogeneous between themselves, but internally homogeneous) that can be identified within the facet (Wright & Masters, 1982). If the cutoff point is set to 3 standard deviation points and standard deviation for measurement error is considered, then $H = (4G + 1)/3$. H is useful for interpreting the participants and stimuli facets and is less useful for the conditions facet. Indeed, H assumes that the estimates are normally distributed (Wright & Masters, 2002). When elements of the facets are too few to run a test of normality, H cannot be computed.

The last statistic of this group is *separation reliability* (R), which indicates how well the elements of a facet separate out to reliably represent the facet. It reflects an estimation of the relationship between true scores and true variance; therefore, $R = (\text{true } SD)^2/(\text{observed } SD)^2 = G^2/(1 + G^2)$, where observed SD is the standard deviation of the estimates (not corrected for measurement error). If $R < .5$, the value of G (separation) is probably due to measurement error. The expected value is high if homogeneity is expected between the facets and low if separation is expected. For example, in a situation in which several judges rate a series of participants on N factors, it would be expected that R is high for factors and for participants but is low for judges, so that we can say that the factors measure the same dimension, that the ratings have discriminated between the various participants, and that the judges are consistent in their rating (Myford & Wolfe, 2003). In the case of experimental procedures for the assessment of automatic associations, the reliability (R) of the items gives us a measure of their equivalence (or interchangeability). Thus, it is desirable to obtain *low reliability indexes for the facet item*.

In our data, the separation index G of the conditions facet is 12.70; hence, the group-level implicit association is very high and reliable ($R = .99$).

Turning to the other facets, participants were centered around a logit value of 0 ($SD = .51$). Among them, we can find 3 outliers below logit −1.00 (least accurate) and one above logit 1.00 (most accurate). Both from a visual inspection of Figure 1 and according to the Kolmogorov–Smirnov test (Chakravarti, Laha, & Roy, 1967), stimuli and participant measures can be considered normally distributed ($Z_{\text{stimuli}} = .789$, $p = .56$; $Z_{\text{part}} = .48$, $p = .98$); hence, the assumptions for the computation of H are likely to be met. The separation ratio (G) of the facet participants was 3.25. The index of separation indicates that the number of heterogeneous groups that can be identified is four (at 3 SDs from one another). The reliability of this separation (R) is .91. These values indicate moderate individual differences in the ability to categorize stimuli in the GNAT. As was already noted, participants' ability on the categorization task is not of direct interest when an implicit measure is analyzed, but low person separations are preferable, because it would suggest that the technique is not influenced by participants' task-set switching ability (see, e.g., McFarland & Crouch, 2002). MFRM estimates, however, are sample independent (a consequence of SO; see the Appendix); hence, even if the technique might suffer from such a potential confound, the Rasch estimates of the implicit associations do not.

The MFRM also provides useful information about the stimuli used to measure the implicit construct under investigation. Low-fitting items are dangerous for implicit techniques because they might be ambiguous or they might represent concepts other than those activated by the other stimuli, whereas extreme items are dangerous because they might trigger reverse-priming effects (Glaser & Banaji, 1999). In this case, the fit indexes for the stimuli facets are all largely inside the acceptable range ($0.94 < \text{infit} < 1.08$), suggesting unidimensionality. Yet, as can be seen in Figure 1, some extreme items (>2 *SD*s from the mean logit) were included in the procedure but should be avoided in future applications. Specifically, they are pictures of mass-produced jelly rolls, which were very different from the homemade cakes represented by the other stimuli. On the other hand, a very attractive chocolate cake (a "Sacher torte") turned out to be too easy to be categorized and should be avoided as well. The reliability of the stimuli is .75 ($G = 1.74$, $H = 2.76$), meaning that the sample is sufficiently large (although not huge) and that the stimuli measures are adequately spread.

The time when measures were taken shows a learning effect that influences the categorization task [$\chi^2(2) = 189.60$, $p < .001$; $G = 7.46$; $R = .98$]. Time 1 is the most difficult, and Times 2 and 3 do not differ with each other [$t(1) = 1.20$, $p = .44$]. The second and third time the participants took the GNAT, they were more accurate. However, this learning effect does not influence the individual estimates of implicit attitude toward sweet food [$F(2,56) = 0.36$, $p = .70$] and salty food [$F(2,56) = 1.01$, $p = .37$].

MFRM also provides the possibility of running bias/interaction analyses between two or more facets. The *differential person functioning* (DPF) analyzes the interaction between the elements of the facet *participants* and elements of other facets. Of particular interest in this context is the interaction with the facet *conditions*. The *bias index* involves introducing an interaction parameter into the model between the facets (e.g., ξ_{nj} for the participants × conditions interaction). With the aim of evaluating interaction significance, such parameters are often transformed into *t* points according to the following formula:

$$t = \frac{\xi_{nj}}{\sqrt{SE_n^2 + SE_j^2}}. \qquad (10)$$

Bias terms (ξ_{nj}) are calculated using a two-stage calibration. In the first stage, the incomplete model, without interaction, is estimated. Subsequently, all the parameters are linked to the values that have been previously calculated, and only ξ_{nj} is estimated (Linacre, 2009a). The bias term represents the distance between expected and observed scores in logit units. In our analysis, it is an estimate of the implicit association between each pair of categories included in the task (i.e., salty+bad, salty+good, sweet+bad, sweet+good). Differential estimates of the implicit attitude toward the target can be obtained by subtracting the value of the beta parameter of a condition (e.g., sweet+bad) from the value of the beta parameter of another condition (e.g., sweet+good). These are also called pairwise contrasts, and their standard error is

$$SE_{nj} = \sqrt{SE_n^2 + SE_j^2},$$

also called *joint SE* (Linacre, 2009a). Dividing the contrast by its joint *SE*, a *t* is obtained.

The plot in Figure 2 provides individual *t* values from the pairwise comparison between the negative and the positive conditions of each target (sweet and salty food). Altogether, 15 participants show significant implicit associations. Six of them, which are highlighted by squares in the figure, show a positive implicit attitude toward sweet food. Five show a positive implicit attitude toward salty food (circles), and 4 show a negative implicit attitude toward salty foods (triangles). As an example, Table 2 provides relevant information for the 9 participants who showed a significant positive or negative implicit evaluation of salty food.

In this table, the "Salty+Good" and "Salty+Bad" columns show the difficulty (in logit) of the conditions (including measurement error). The "Contrast" column displays the difference in logit between the two conditions, the standard error of contrast, the value of *t* associated with such a difference, the *df*s, and the associated significance levels (two-tailed).

The *t* values that can be computed with a DPF analysis are standardized, usually reliable, easily interpretable, and normally distributed (if $df > 30$) estimates of the individual implicit association studied. They can be computed both for a single element (dividing the Rasch measure by its *SE*) and for a difference of logits (using the joint *SE*). The *reliability* (r_{00}) of these MFRM-derived implicit association scores can be computed according to its classical definition (true variance + error variance = observed variance). Specifically, the variance of the single (ξ_{nj}) or pairwise (*contrast*) measure of association across participants ("true" variance) should be divided by the sum of true variance and error variance (the mean across participants of the squared standard errors). Table 3 provides mean Rasch measures of association, the standard deviations, and the reliabilities of these measures, separately by the time in which the GNAT was taken. As can be seen in Table 4, these estimates are substantially correlated with the scoring procedure originally suggested for the GNAT (d'), but they never share more than 38% of variance. The d' statistic is computed as the difference between the standardized proportion of correct responses to targets (hits) and the standardized proportion of errors to distractors (false alarms). According to signal detection theory (Green & Swets, 1966), this index provides a measure of the individual ability to discriminate signals (target stimuli) with noise (distractors) from noise alone. Although the lack of correlation we found between the two alternate scoring methods might be due to the low reliability of d' scores

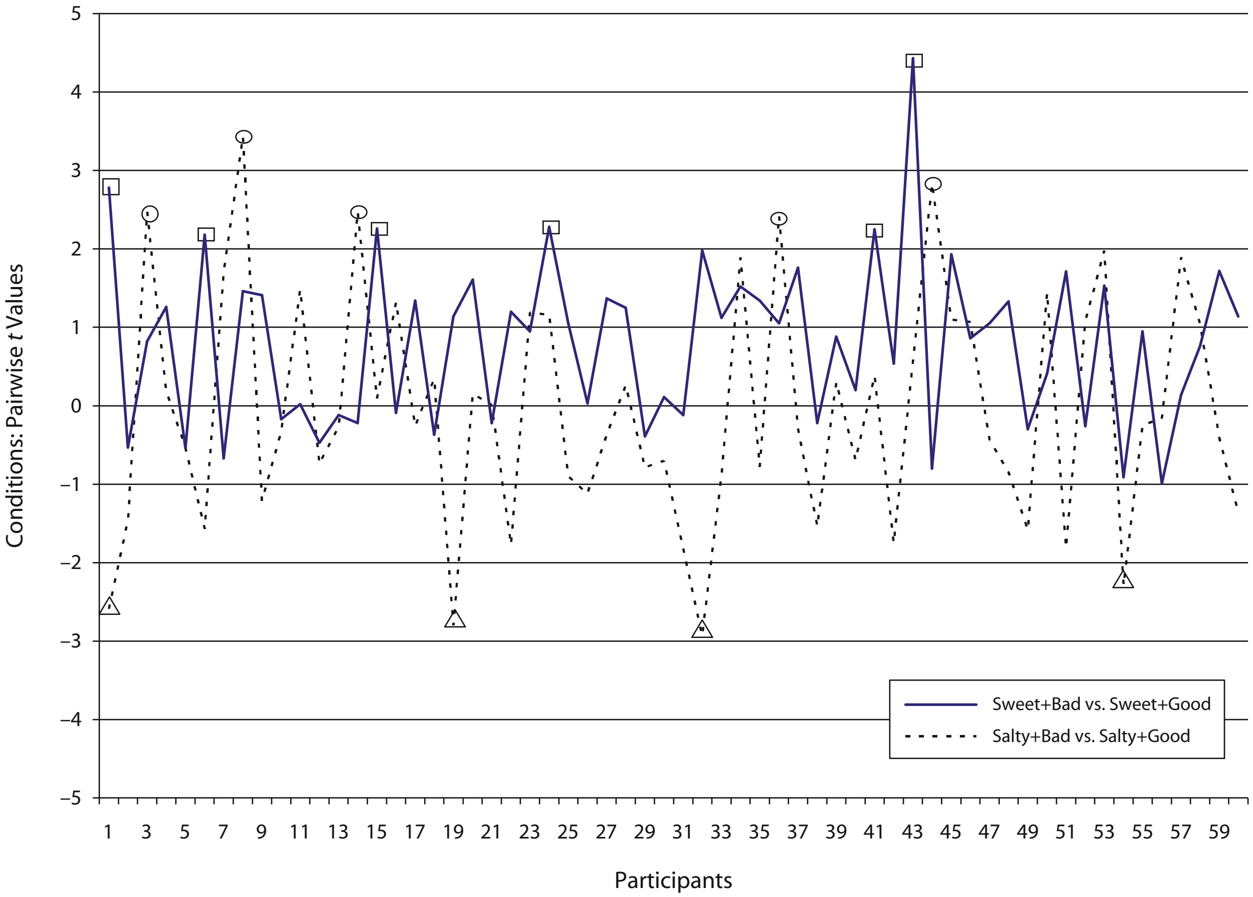

Figure 2. Plot of individual estimates (*t*) of the implicit attitude toward sweet and salty food. Positive *t* values indicate positive implicit attitude; *t* values higher than |2| indicate a significant implicit association.

Table 2
Differential Person Functioning Analysis of 9 Participants Who Showed Significant Implicit Associations Toward Sweet or Salty Foods

	Salty+Bad		Salty+Good		Contrast (Bad–Good)				
Participant	Measure	*SE*	Measure	*SE*	Measure	*SE*	*t*	*df*	*p*(*t*)
8	−0.15	0.24	−1.28	0.22	1.13	0.33	1.64	219	<.001
3	0.59	0.31	−0.45	0.29	1.04	0.42	−1.86	222	.014
36	0.39	0.29	−0.57	0.27	0.96	0.40	−1.35	219	.016
44	−0.54	0.22	−1.41	0.21	0.87	0.31	−1.38	219	.005
14	−0.30	0.23	−1.11	0.23	0.81	0.33	−1.38	219	.014
1	−0.88	0.21	0.21	0.37	−1.09	0.42	−1.51	203	.010
54	−0.34	0.23	0.94	0.51	−1.28	0.56	−1.81	192	.023
19	−0.76	0.21	0.56	0.42	−1.32	0.47	1.77	197	.005
32	−0.76	0.21	0.72	0.46	−1.48	0.50	−2.24	192	.003

(mean $\alpha = .22$), we think that the two measures are actually different and that Rasch estimates are preferable, because they are more reliable (mean $r_{00} = .58$). A direct comparison against concurrent measures (e.g., other implicit and/or explicit measures of the same construct or measures of actual behavior theoretically related to the construct under investigation) might indicate which model, and related scoring, better fits the needs and expectations of a researcher using the GNAT or any other implicit technique.

DISCUSSION

The present study has demonstrated the applicability of the MFRM to data obtained using the GNAT procedure and the multiple ways in which it can be used. The most relevant aspects will now be discussed.

Latent Trait

The statistical procedure allows the determination of the goodness of fit of the data to the model through a

Table 3
Means, Standard Deviations, and Reliabilities of the Rasch Measures of Implicit Association

Implicit Association	Time	*M*	*SD*	Reliability
Sweet+Bad	1	0.14	0.83	.58
	2	0.31	0.91	.50
	3	0.33	0.97	.51
Sweet+Good	1	0.14	0.50	.51
	2	0.01	0.75	.60
	3	0.18	0.86	.61
Salty+Bad	1	0.17	0.71	.64
	2	0.16	0.85	.67
	3	0.12	0.70	.58
Salty+Good	1	0.25	0.89	.61
	2	0.25	0.86	.46
	3	0.31	0.92	.48
Sweet (Bad vs. Good)	1	0.01	0.96	.57
	2	0.30	1.18	.50
	3	0.03	1.20	.55
Salty (Bad vs. Good)	1	−0.08	1.06	.63
	2	−0.09	1.22	.62
	3	−0.22	1.05	.55

Note—Values in the third column are mean ξ_{nj} for the associations Sweet+Bad, Sweet+Good, Salty+Bad, and Salty+Good and mean contrasts for the associations Sweet (Bad vs. Good) and Salty (Bad vs. Good). The reliability is computed according to its classical definition (true variance + error variance = observed variance). Specifically, we divided the variance of the single (ξ_{nj}) or pairwise (contrast) measure of association across participants ("true" variance) by the sum of true variance and error variance (the mean across participants of the squared standard errors).

specific index of global fit (L^2) and through item-level standardized infit and outfit statistics that, if adequate, allow us to consider the facets as aspects of the same trait, with a common measurement unit. Our results showed that all standardized infit and outfit statistics were far away from the critical level of 2. Furthermore, we showed that participants' parameters do not change whether the positive or the negative stimuli alone are included in the analysis and that a principal components analysis on standardized residuals extracts components with an eigenvalue smaller than 3 (see Linacre, 2009b). Hence, we can say that a unidimensional latent trait has been defined as precision on the association task, for participants, items, times of observation, and conditions of association.

Reliability Analysis

Specific indices (*R*, *G*, and *H*) allow the analysis of the reliability of the separation of the elements in each facet. In particular, the person separation index ($H = 4.66$) has shown that the four groups for which the task is differently difficult can at least be identified. The item reliability index ($R = .75$) showed that the sample size is adequate, although not huge. The separation ratio of the time facet ($G = 7.46$) showed that a learning effect took place, since participants are much more accurate in Times 2 and 3. The reliability index for the last facet suggests that the conditions are reliably different ($R = .99$) and indicates the existence of a group-level implicit positive attitude toward salty food and a negative implicit attitude toward sweet food.

Item Analysis

The infit and outfit values indicate the goodness of fit of the items to the latent trait and allow the selection of the more unidimensional stimuli. Furthermore, this analysis can be extended to any other facet of the model, such as, for example, the conditions or the participants. The estimates of the parameters and the associated standard errors allow the statistical evaluation of the different/identical location of the stimuli on the latent trait and, therefore, their recognizability. In implicit techniques, the choice of stimuli is extremely important, because they directly affect the validity of the measure. They should adequately represent the subject of the study, and they should all be equally recognizable. When response times or recognition errors of a categorization task are analyzed, the δ parameters, which typically reflect item difficulty, represent both their *recognizability* and their *prototypicality*, as compared with the nominal category of interest. The length of a word can mean that more time is needed to read it and, consequently, more time to respond. But even the prototypicality of a stimulus can influence the accuracy (or the speed) of the response, both in reading and in making decisions. For example, the term *bat*, although easy to read, is not especially representative of the category *mammals*, which would be better represented by a stimulus such as *monkey* or *elephant*. The analysis of the δ parameters allows the diagnosis of any anomalies arising from the choice of stimuli.

Interactions

The model allows an analysis of the interaction between various facets. We considered interactions between participants and conditions (differential person functioning), because they provide individual estimates of the implicit associations under investigation and represent a scoring procedure that is different from that already present in the literature (d'; Nosek & Banaji, 2001). Although we have seen that our Rasch-based estimates of implicit association are much more reliable than d' scores, future criterion-related validity studies might suggest which

Table 4
Zero-Order Pearson Correlations Among MFRM Standardized Estimates of Individual Associations (*t*) and *d′* Scores

Association	Sweet+Bad (*t*)	Sweet+Good (*t*)	Salty+Bad (*t*)	Salty+Good (*t*)
Sweet+Bad (*d′*)	$.560^{**}$	−.120	$-.349^{**}$	−.002
Sweet+Good (*d′*)	−.144	$.499^{**}$	−.178	−.184
Salty+Bad (*d′*)	−.240	−.179	$.498^{**}$	−.148
Salty+Good (*d′*)	−.034	$-.255^{*}$	$-.299^{*}$	$.616^{**}$

$^{*}p < .05$. $^{**}p < .01$.

scoring best suits the needs of an implicit measurement. However, from a mathematical point of view, *t* values derived from a Rasch analysis are superior to *d′* for many reasons. First, the *d′* cannot be computed for participants with an error rate greater than 50% or with a proportion of hits and false alarms equal to 0% or 100%. Then, *t* values are preferable because we know their distribution and because they are computed from measures that are independent, by definition, from the elements of all the other facets included in the analysis. As a consequence, there is no risk that a Rasch-based measure of implicit association is influenced by participants' ability or by the time when measures are taken, which have been found to be serious potential confounds (McFarland & Crouch, 2002; Robusto, Cristante, & Vianello, 2008). Lastly, Blanton and Jaccard (2006) recently raised an issue against the arbitrariness of implicit measures. The MFRM elegantly solves this potential problem, because it attributes a rational zero point to the measurement undertaken, which can be fixed to the mean of the implicit measure obtained.

CONCLUSIONS

Implicit measures have had a great impact in psychological research and have been extensively studied. At the date this article was written, the three most used techniques (EP, Fazio et al., 1986; IAT, Greenwald et al., 1998; GNAT, Nosek & Banaji, 2001) had been cited in almost 3,500 different articles (source: Google Scholar). Yet they are far from being a closed chapter, from both a theoretical and a methodological point of view. For the theory, many contributions keep clarifying what an implicit measure is (and is not). For instance, Gawronski, LeBel, and Peters (2007) recently questioned some common assumptions about the stability, the lack of consciousness, and the resistance to social desirability of implicit measures, concluding that the available evidence is still equivocal. With respect to the method, a number of articles have scrutinized their psychometric properties, which have also been recently and comprehensively reviewed or meta-analyzed in many others (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Hofmann & Schmitt, 2008; Lane, Banaji, Nosek, & Greenwald, 2007; Schnabel, Asendorpf, & Greenwald, 2008). In addition, some contributions have proposed empirically validated scoring algorithms (Greenwald, Nosek, & Banaji, 2003), multinomial models to separate automatic and controlled processes (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005), and applications of formal models previously developed to analyze latencies (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007). Yet many methodological issues are still open. For example, no previous contribution has successfully solved the problems of metric arbitrariness (Blanton & Jaccard, 2006) and those related to participants' ability (McFarland & Crouch, 2002). These issues have been addressed in this article, in which the MFRM was proposed as a formal model for the analysis of the GNAT. Specifically, we proposed an analysis strategy that fits implicit measures and the most common needs of the researchers that use them, showing that the MFRM provides a number of statistics and useful information that is not available in other models. For example, we described the separation index *G*, which gives a sample-level effect size of the intensity of the implicit associations under investigation; the logit scale and the fit indexes, according to which bad stimuli can be avoided in future applications; and finally, the bias term and the *t* value associated as individual estimates of the implicit associations.

Notably, we found that Rasch-based individual measures of implicit associations are much more reliable than the *d′* scores originally proposed for the GNAT. Furthermore, among the many benefits, the MFRM resolves the issue of arbitrariness (Blanton & Jaccard, 2006), because MFRM estimates are centered by construction around a rational zero point (e.g., their mean). Lastly, Rasch-based individual measures of implicit association are, by definition, independent from participants' task-set switching ability; hence, they also prevent the implicit measure from being affected by a potential confound that was first studied by McFarland and Crouch (2002).

The measurement model and the analysis strategy we adopted in this article should easily fit other implicit techniques, such as the AMP (Payne et al., 2005). Yet some more formal research has to be conducted before a many-facet Rasch model for continuous variables can be used to analyze latency-based techniques of implicit constructs, such as the IAT. In the meanwhile, we hope that this article will motivate researchers to use the MFRM to analyze errors (e.g., in a GNAT) and dichotomous choices (e.g., in an AMP), in order to construct high-quality and readily useful measures of the implicit constructs they are investigating.

AUTHOR NOTE

Correspondence concerning this article should be addressed to M. Vianello (e-mail: michelangelo.vianello@unipd.it).

REFERENCES

Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills, CA: Sage.

Banaji, M., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality & Social Psychology*, **68**, 181-198.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology*, **56**, 329-343.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, **61**, 27-41.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measures of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality & Social Psychology*, **79**, 631-643.

Chakravarti, I. M., Laha, R. G., & Roy, J. (1967). *Handbook of methods of applied statistics: Vol. I*. New York: Wiley.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality & Social Psychology*, **89**, 469-487.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, **50**, 77-85.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, **54**, 297-327.

FAZIO, R. H., SANBONMATSU, D. M., POWELL, M. C., & KARDES, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality & Social Psychology*, **50**, 229-238.

FISCHER, G. H. (1995). Derivations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications*. New York: Springer.

FISHER, R. A. (1970). *Statistical methods for research workers* (14th ed.). Edinburgh: Oliver & Boyd.

GAWRONSKI, B., LEBEL, E. P., & PETERS, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, **2**, 181-193.

GLASER, J., & BANAJI, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality & Social Psychology*, **77**, 669-687.

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

GREENWALD, A. G., MCGHEE, D. E., & SCHWARTZ, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality & Social Psychology*, **74**, 1464-1480.

GREENWALD, A. G., NOSEK, B. A., & BANAJI, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality & Social Psychology*, **85**, 197-216.

GREENWALD, A. G., POEHLMAN, T. A., UHLMANN, E., & BANAJI, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality & Social Psychology*, **97**, 17-41.

HIGGINS, E. T. (1996). Knowledge activation: Accessibility, applicability and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford.

HOFMANN, W., & SCHMITT, M. (2008). Advances and challenges in the indirect measurement of individual differences at age 10 of the Implicit Association Test. *European Journal of Psychological Assessment*, **24**, 207-209.

KARPINSKI, A., & STEINMEN, R. B. (2006). The single category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality & Social Psychology*, **91**, 16-32.

KLAUER, K. C., VOSS, A., SCHMITZ, F., & TEIGE-MOCIGEMBA, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality & Social Psychology*, **93**, 353-368.

LANE, K. A., BANAJI, M. R., NOSEK, B. A., & GREENWALD, A. G. (2007). Understanding and using the Implicit Association Test: IV. Procedures and validity. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59-102). New York: Guilford.

LINACRE, J. M. (1989). *Multi-facet Rasch measurement*. Chicago: MESA Press.

LINACRE, J. M. (2009a). *Facets Rasch measurement computer program*. Chicago: Winsteps.com.

LINACRE, J. M. (2009b). *A user's guide to WINSTEPS/MINISTEP Rasch-model computer programs*. Chicago: Winsteps.com.

LINACRE, J. M., & WRIGHT, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, **8**, 370. Available at http://rasch.org/rmt/rmt83.htm.

MCFARLAND, S. G., & CROUCH, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition*, **20**, 483-510.

MYFORD, C. M., & WOLFE, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, **4**, 386-422.

NOSEK, B. A., & BANAJI, M. R. (2001). The go/no-go association task. *Social Cognition*, **19**, 625-666.

NOSEK, B. A., GREENWALD, A. G., & BANAJI, M. R. (2006). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265-292). Philadelphia: Psychology Press.

NUNNALLY, J. C., & BERNSTEIN, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

OLSON, M. A., & FAZIO, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, **14**, 636-639.

PAYNE, B. K., CHENG, C. M., GOVORUN, O., & STEWART, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality & Social Psychology*, **89**, 277-293.

RASCH, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. (Original work published 1960)

ROBUSTO, E., CRISTANTE, F., & VIANELLO, M. (2008). Assessing the impact of replication on Implicit Association Test effects by means of the extended logistic model for the assessment of change. *Behavior Research Methods*, **40**, 954-960.

SCHNABEL, K., ASENDORPF, J. B., & GREENWALD, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment*, **24**, 210-217.

SMITH, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, **3**, 25-40.

SMITH, R. M. (2002). Detecting and evaluating the impact of multidimensionality using item statistics and principal component analysis of residuals. *Journal of Applied Measurement*, **3**, 205-231.

WRIGHT, B. D., & MASTERS, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

WRIGHT, B. D., & MASTERS, G. N. (2002). Number of person or item strata. *Rasch Measurement Transactions*, **16**, 888.

APPENDIX
Specific Objectivity in Rasch Measurement

Rasch's favorite example to explain the meaning of specific objectivity (Rasch, 1960/1980) was related to the joint definition and measurement of mass and force in classical mechanics (see Fischer, 1995). Let O_v, $v = 1, 2, \ldots,$ be rigid bodies, and let their masses be M_v. Furthermore, let there be some experimental conditions in which forces F_i are applied to each of the masses, such as to produce acceleration A_{vi}. According to the second Newtonian axiom (force = mass × acceleration), acceleration is proportional to the force exerted on the object and is inversely proportional to the object's mass, $A_{vi} = M_v^{-1} F_i$. Therefore, any two masses M_v and M_w can be compared according to the quotient

$$\frac{A_v}{A_w} = \frac{M_v^{-1} F_i}{M_w^{-1} F_i} = \frac{M_v}{M_w}.$$

This implies that the two masses can be compared independently of (i.e., without knowledge of) the forces that are applied.

Rasch (1960/1980) developed a model in which these specific objective comparisons can be applied to participants and items of a test.

The demonstration follows considering that the probability of obtaining a given pattern of responses x from a participant n to k items is the product of the probabilities that the participant n has of giving each response:

$$P\left(x_{n1}, x_{n2}, \ldots, x_{nk}\right) = P\left(x_{n1}\right) P\left(x_{n2}\right) \ldots P\left(x_{nk}\right). \tag{A1}$$

Equation 1 describes the probability of a response $X_{ni} = (1, 0)$,

$$P(X_{ni} = x_{ni} \mid \beta_n, \delta_i) = \frac{\exp\left[x_{ni}\left(\beta_n - \delta_i\right)\right]}{1 + \exp\left(\beta_n - \delta_i\right)};$$

hence, Equation A1 can be written as

$$\begin{aligned} P\left(x_{n1}, x_{n2}, \ldots, x_{ni}\right) &= \frac{\exp\left[x_{n1}\left(\beta_n - \delta_1\right)\right] \exp\left[x_{n2}\left(\beta_n - \delta_2\right)\right] \cdots \exp\left[x_{ni}\left(\beta_n - \delta_i\right)\right]}{\left[1 + \exp\left(\beta_n - \delta_1\right)\right]\left[1 + \exp\left(\beta_n - \delta_2\right)\right] \cdots \left[1 + \exp\left(\beta_n - \delta_i\right)\right]} \\ &= \frac{\exp\left(x_{n1}\beta_n - x_{n1}\delta_1 + x_{n2}\beta_n - x_{n2}\delta_2 + \cdots + x_{nk}\beta_n - x_{nk}\delta_k\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_v - \delta_i\right)} \\ &= \frac{\exp\left[\beta_n\left(x_{n1} + x_{n2} + \cdots + x_{nk}\right) - x_{n1}\delta_1 - x_{n2}\delta_2 + \cdots - x_{nk}\delta_k\right]}{\prod_{i=1}^{k} 1 + \exp\left(\beta_v - \delta_i\right)}. \end{aligned}$$

Now, given that

$$\left(x_{n1} + x_{n2} + \cdots + x_{nk}\right) = \sum_{i=1}^{k} x_{ni} = r_n$$

represents the pattern of responses of participant n to k items, we can synthesize, writing

$$P\left(x_{n1}, x_{n2}, \ldots, x_{nk}\right) = \frac{\exp\left(\beta_n r_n - \sum_{i=1}^{k} x_{ni}\delta_i\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}. \tag{A2}$$

Consider two items; if the sum of their scores is $r_n = 1$, the possible patterns are (1, 0) and (0, 1), and the probabilities of these, following Equation A2, are

$$P(1,0) = \frac{\exp\left(\beta_n 1 - 1\delta_1 - 0\delta_2\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)} = \frac{\exp\left(\beta_n - \delta_1\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}$$

and

$$P(0,1) = \frac{\exp\left(\beta_n 1 - 0\delta_1 - 1\delta_2\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)} = \frac{\exp\left(\beta_n - \delta_1\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}.$$

APPENDIX (Continued)

As a consequence, the probability of a given r_n is the sum of the probabilities of all possible patterns that can produce r_n:

$$P\left(r_n = x_{n1} + x_{n2} + \cdots + x_{nk}\right) = \sum_{x_{n1}, x_{n2}, \ldots, x_{nk} \mid r_n} \frac{\exp\left(\beta_n r_n - \sum_{i=1}^{k} x_{ni}\delta_i\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}. \tag{A3}$$

The conditional probability of a pattern given r_n is the ratio between the probability of that pattern and the probability of obtaining any other pattern with the same r_n.

Given two items and $r_n = 1$, the probability of a specific pattern—(1, 0), for example—is

$$P\left[(1,0) \mid r_n = 1\right] = \frac{P(1,0)}{P(1,0) + P(0,1)},$$

in which the denominator is justified because, with dichotomies, two cases exist for which $r_n = 1$: (1, 0) and (0, 1).

The conditional probability of a pattern given r_n is therefore obtained by dividing Equation A2 by Equation A3:

$$P\left(x_{n1}, x_{n2}, \ldots, x_{nk} \mid r_n\right) = \frac{\dfrac{\exp\left(\beta_n r_n - \sum_{i=1}^{k} x_{ni}\delta_i\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}}{\sum_{x_{n1}, x_{n2}, \ldots, x_{nk} \mid r_n} \dfrac{\exp\left(\beta_n r_n - \sum_{i=1}^{k} x_{ni}\delta_i\right)}{\prod_{i=1}^{k} 1 + \exp\left(\beta_n - \delta_i\right)}}$$

$$= \frac{\exp\left(\beta_n r_n - x_{n1}\delta_1 - x_{n2}\delta_2 + \cdots - x_{nk}\delta_k\right)}{\exp\left(\beta_n r_n - x_{n1}\delta_1\right) + \exp\left(\beta_n r_n - x_{n2}\delta_2\right) + \cdots + \exp\left(\beta_n r_n - x_{nk}\delta_k\right)}$$

$$= \frac{\exp\left(\beta_n r_n\right)\exp\left[-\left(x_{n1}\delta_1\right)\right]\exp\left[-\left(x_{n2}\delta_2\right)\right]\cdots\exp\left[-\left(x_{nk}\delta_k\right)\right]}{\exp\left(\beta_n r_n\right)\exp\left[-\left(x_{n1}\delta_1\right)\right] + \exp\left(\beta_n r_n\right)\exp\left[-\left(x_{n2}\delta_2\right)\right] + \cdots + \exp\left(\beta_n r_n\right)\exp\left[-\left(x_{nk}\delta_k\right)\right]}$$

$$= \frac{\exp\left(\beta_n r_n\right)\exp\left[-\left(x_{n1}\delta_1\right)\right]\exp\left[-\left(x_{n2}\delta_2\right)\right]\cdots\exp\left[-\left(x_{nk}\delta_k\right)\right]}{\exp\left(\beta_n r_n\right)\left\{\exp\left[-\left(x_{n1}\delta_1\right)\right] + \exp\left[-\left(x_{n2}\delta_2\right)\right] + \cdots + \exp\left[-\left(x_{nk}\delta_k\right)\right]\right\}}$$

$$= \frac{\exp\left[-\left(x_{n1}\delta_1\right)\right]\exp\left[-\left(x_{n2}\delta_2\right)\right]\cdots\exp\left[-\left(x_{nk}\delta_k\right)\right]}{\exp\left[-\left(x_{n1}\delta_1\right)\right] + \exp\left[-\left(x_{n2}\delta_2\right)\right] + \cdots + \exp\left[-\left(x_{nk}\delta_k\right)\right]}$$

$$= \frac{\exp\left(-\sum_{i=1}^{k} x_{ni}\delta_i\right)}{\sum_{(x_{n1}, x_{n2}, \ldots, x_{nk}) \mid r_n} \exp\left(-\sum_{i=1}^{k} x_{ni}\delta_i\right)}. \tag{A4}$$

Note that β_n is not present in Equation A4 anymore. Hence, it is evident that

- – the distribution of probabilities of r_n is only a function of items' difficulties (δ); furthermore, since the pattern of responses of each individual does not contain more information about those provided by r_n, this is considered a sufficient statistic to estimate β_n;
- – the estimation of items' difficulties is completely independent from participants' ability.

This demonstration is analogous for each facet of the Rasch model.

(Manuscript received July 24, 2009;
revision accepted for publication April 3, 2010.)