Mixing the implicit: A Linear Mixed-Effects Models approach for a Rasch analysis of the Implicit Association Test and the Single Category Implicit Association Test

Ottavia M. Epifania, Pasquale Anselmi, and Egidio Robusto

September 10, 2021

Abstract

The indirect investigation of psychological constructs has become prominent in social sciences thanks to the so-called implicit measures. Different implicit measures can be administered concurrently to the same respondents for obtaining detailed and multifaceted information on the constructs of interest. In this study, a Rasch analysis of accuracy and time responses of two commonly used implicit measures is presented. The focus in on the concurrent administration of the Implicit Association Test (IAT; Greenwald et al., 1998) and the Single Category IAT (SC-IAT; Karpinski & Steinman, 2006). Linear Mixed-Effects Models are used to address the within- and between-measures sources of variability and to obtain a Rasch parametrization of the data. By disentangling the respondent’s contribution from the stimulus contribution to the observed responses, these models allow for delving deeper on the functioning of the IAT and the SC-IAT, as well as for grasping a better understanding of the processes driving a behavioral decision. Implications of the results for social sciences and future research directions are discussed.
Mixing the implicit: A Linear Mixed-Effects Models approach for a Rasch analysis of the Implicit Association Test and the Single Category Implicit Association Test

Ottavia M. Epifania, Egidio Robusto, and Pasquale Anselmi
Department of Philosophy, Sociology, Education, and Applied Psychology,
University of Padova

Author Note

Ottavia M. Epifania  
https://orcid.org/0000-0001-8552-568X

Egidio Robusto  
https://orcid.org/0000-0002-7583-2587

Pasquale Anselmi  
https://orcid.org/0000-0003-2982-7178

The authors have no conflict of interest to disclose.

Authors would like to thank Dr. Paul De Boeck for his suggestions on a previous version of the manuscript.

Correspondence concerning this Article should be addressed to: Ottavia M. Epifania, Department of Philosophy, Sociology, Education, and Applied Psychology, Via Venezia 14, Padova, Italy. E-Mail: ottavia.epifania@unipd.it
Abstract

The indirect investigation of psychological constructs has become prominent in social sciences thanks to the so-called implicit measures. Different implicit measures can be administered concurrently to the same respondents for obtaining detailed and multifaceted information on the constructs of interest. In this study, a Rasch analysis of accuracy and time responses of two commonly used implicit measures is presented. The focus is on the concurrent administration of the Implicit Association Test (IAT; Greenwald et al., 1998) and the Single Category IAT (SC-IAT; Karpinski & Steinman, 2006). Linear Mixed-Effects Models are used to address the within- and between-measures sources of variability and to obtain a Rasch parametrization of the data. By disentangling the respondent’s contribution from the stimulus contribution to the observed responses, these models allow for delving deeper on the functioning of the IAT and the SC-IAT, as well as for grasping a better understanding of the processes driving a behavioral decision. Implications of the results for social sciences and future research directions are discussed.

**Keywords:** Rasch Model, Log-normal Model, Implicit Association Test, Single Category Implicit Association Test, Implicit social cognition.
Mixing the implicit: A Linear Mixed-Effects Models approach for a Rasch analysis of the Implicit Association Test and the Single Category Implicit Association Test

The label “implicit social cognition” identifies a field of investigation where attitudes, opinions, and preferences of respondents are indirectly inferred from processes that are automatically activated by triggering stimuli (i.e., automatic processes). The so-called implicit measures have been introduced to assess these processes through the performance of the respondents at speeded categorization tasks (Greenwald & Lai, 2020). Many implicit measures are available, each providing unique information on the construct. The main difference concerns the indirect technique on which implicit measures rely, namely priming (i.e., the prior presentation of a stimulus –prime– is supposed to affect the evaluation of a second stimulus –target–) or automatic associations (i.e., the exemplars belonging to categories strongly associated between each other are supposed to be more efficiently categorized together than exemplars of categories loosely associated together). The Affective Misattribution Procedure (AMP; Payne et al., 2005) and the Evaluative Priming Task (EPT; Fazio et al., 1986) are measures based on priming. Implicit measures can be further differentiated according to the task (i.e., go/no-go procedure vs. two-choice task, see Gomez et al., 2007), and the measure they provide (i.e., comparative between two contrasting objects vs. “absolute” towards one object). The Implicit Association Test (Greenwald et al., 1998), the Single Category IAT (SC-IAT Karpinski & Steinman, 2006), and the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001) are based on automatic associations. The former two measures are two-choice tasks (i.e., responses are expressed through two distinct response keys), while the latter one is based on a go/no-go procedure (i.e., responses are expressed through a unique response key). The IAT provides a comparative measure between two contrasting targets, while both SC-IAT and GNAT provide an “absolute” measure of one target. Implicit measures can be administered together to obtain
multiple indirect assessments of the same construct, this allowing for deeper understanding of the variables of interest. For instance, the IAT has been administered with the AMP (e.g., [Green et al., 2019]), with the SC-IAT, (e.g., [Richetin et al., 2019]), and with the GNAT (Ueda et al., 2017). The AMP has been administered with the SC-IAT (e.g., [Richard et al., 2017]) and with the Brief IAT (B-IAT Sriram & Greenwald, 2009) in Miles, Charron-Chénier, and Schleifer (2019). When implicit measures are administered together in a within-subjects experimental design, sources of random variability due to the within–respondents variability are expected, and they should be accounted for to obtain meaningful estimates. In this contribution, a Rasch modeling based on Linear Mixed-Effects Models (LMMs) is presented to obtain meaningful information on both respondent’s performance and stimulus functioning while addressing the sources of variability in the data. The focus is on the concurrent administration of the IAT and the SC-IAT, which are two of the most common implicit measures (Epifania, Robusto, & Anselmi, 2020a). However, it is worth noting that this framework is easily extensible for modeling data of other implicit measures administered together.

The IAT and the SC-IAT assess the strength of the associations between targets (e.g., Coke and Pepsi in a Soda IAT, Coke in a Coke SC-IAT) and evaluative dimensions (Good and Bad) by measuring the speed and accuracy with which prototypical exemplars presented on the computer screen are sorted in their own category with two response keys. In both measures, the categorization task takes place in two contrasting conditions. In one associative condition of the IAT (i.e., Coke-Good/Pepsi-Bad condition), Coke and Good exemplars are assigned to their categories with the same response key, while Pepsi and Bad exemplars are assigned with the opposite key. In the contrasting condition (i.e., Pepsi-Good/Coke-Bad condition), Pepsi and Bad exemplars are assigned with the same response key, while Coke and Bad exemplars are assigned with the opposite key. The categorization task in the SC-IAT is almost identical. In one
associative condition of the SC-IAT (i.e., Coke-Good condition), Coke and Good exemplars are assigned with the same response key, while Bad exemplars are assigned with the opposite key. In the contrasting condition (i.e., Coke-Bad condition), Coke and Bad exemplars are assigned with the same response key, while Good exemplars are assigned with the opposite key. The IAT effect$^1$ is the difference in the performance of the respondents between the associative conditions. The IAT effect expresses either how much one target is positively valued in respect to its opposite (IAT) or how much the target is positively or negatively valued (SC-IAT). Consequently, the IAT is useful when the aim is to obtain a comparative measure between two contrasting targets. Conversely, the SC-IAT is particularly useful when a clear contrasted target cannot be found or when the indirect assessment is aimed at absolute evaluations of one target (see, e.g., Karpinski & Steinman 2006). Usually, the strength and direction of the IAT effect are expressed by ad hoc effect size measures, the so-called $D$ scores. The $D$ scores result from the difference in the average response time between conditions divided by the standard deviation computed on the pooled trials of both conditions (Greenwald et al. 2003; Karpinski & Steinman 2006).

The presentation of the same set of stimuli within and between associative conditions to the same respondent(s) (i.e., fully-crossed design, Westfall et al. 2014) generate dependencies between single observations. If the fully-crossed structures and their related sources of variability are not appropriately accounted for, biased estimates of the construct under investigation are obtained, the importance of experimental effects is underestimated, and the probability of committing Type I Error is inflated (Barr et al. 2013; Judd et al. 2012, 2017; McCullagh & Nelder 1989; Westfall et al. 2014; Wolsiefer et al. 2017). When multiple implicit measures are administered together to the same sample of respondents, further sources of dependency due to the within—respondents between—measures variability are expected. Additionally, the use of the

$^1$The term IAT effect is here used to express the bias due to the SC-IAT associative conditions as well.
same set of stimuli across measures generates between measures sources of dependencies at the stimulus level as well. If the data of the implicit measures administered together are analyzed independently from one another, the between–measures sources of variability at respondent and stimulus levels are left free to bias the results. Averaging across trials in each associative condition, the $D$ scores can account for neither the within–measure nor the between–measures sources of variability (Epifania, Robusto, & Anselmi [2020b], Wolsiefer et al., [2017], Westfall et al., [2014]).

In Epifania, Robusto, and Anselmi (2020b), a modeling framework based on Linear Mixed-Effects Models (LMMs) has been introduced to address the fully-crossed design of the IAT data while providing Rasch (Rasch, 1960) and log-normal model estimates (van der Linden, 2006) from accuracy and log-time responses. In this contribution, the modeling framework in Epifania, Robusto, and Anselmi (2020b) is extended to address the sources of variability due to the administration of multiple implicit measures in within-subjects experimental designs. The variability due to the multiple presentations of stimuli across measures is addressed as well. The application of the modeling framework to the specific case of the IAT administered together with the SC-IAT is presented. Nonetheless, this modeling framework is easily extensible to the cases of other implicit measures administered together.

**Rasch, log-normal, and (generalized) linear models**

According to the Rasch model (Rasch, 1960, Equation 1 in Table 1), the probability of respondent $p$ to give a correct response to stimulus $s$ depends on both respondent’s (i.e., ability parameter $\theta_p$, expressing the amount of latent trait possessed by respondent $p$) and stimulus (i.e., difficulty parameter $b_s$, expressing the amount of latent trait needed for giving the correct response to stimulus $s$) characteristics. The higher the value of $\theta_p$, the higher the ability of
respondent \( p \), (i.e., the higher the number of items correctly endorsed by respondent \( p \)). The higher the value of \( b_s \), the higher the difficulty of item \( s \) (i.e., the higher the number of incorrect responses observed on stimulus \( s \)). In a Generalized Linear Model (GLM) for binomially distributed responses, the probability \( \mu_{ps} \) of respondent \( p \) to give a correct response to stimulus \( s \) is yielded by the inverse of the logit link function (i.e., \( \text{logit}^{-1} \), Equation 2 in Table 1). The logit link function is used to link the observed accuracy responses with the linear combination of predictors \( \eta_{ps} \). The Rasch model (Equation 1) can be equated to the inverse of the logit link function \( \text{logit}^{-1} \) of the GLM. Thus, the estimates of the former model can be obtained from the latter one (De Boeck et al., 2011; Doran et al., 2007; Gelman & Hill, 2007).

**Table 1**

<table>
<thead>
<tr>
<th>Rasch, log-normal, and linear models.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Typical</strong></td>
</tr>
<tr>
<td><strong>(G)LM</strong></td>
</tr>
</tbody>
</table>

Rasch model

\[
P(x_{ps}|\theta_p, \delta_s) = \frac{\exp(\theta_p - b_s)}{1 + \exp(\theta_p - b_s)} \tag{1}
\]

\[
\mu_{ps} = \text{logit}^{-1}(\eta_{ps}) = \frac{\exp(\theta_p + b_s)}{1 + \exp(\theta_p + b_s)} \tag{2}
\]

Log-normal model

\[
E(t_{ps}|\tau_p, \delta_s) = \delta_s - \tau_p + \varepsilon \tag{3}
\]

\[
E(t_{ps}|\tau_p, \delta_s) = \delta_s + \tau_p + \varepsilon \tag{4}
\]

**Note:** \( \mu_{ps} \): Probability of a correct response for respondent \( p \) on stimulus \( s \) given the linear combination of predictors \( \eta_{ps} \), \( \eta_{ps} \): Linear combination of predictors \( \theta_p + b_s \), \( \text{logit}^{-1} \): Inverse of the logit link function \( \text{logit} = \log \left( \frac{\mu_{ps}}{1 - \mu_{ps}} \right) \).

The log-normal model (van der Linden, 2006, Equation 3 in Table 1) allows for obtaining a Rasch parametrization of log-time responses by considering the observed log-time response as a function of respondent speed \( \tau_p \) (i.e., the larger the value of speed \( \tau_p \), the smaller the amount of time respondent \( s \) needs for performing the task), and stimulus time intensity \( \delta_s \) (i.e., the
larger the value of time intensity $\delta_s$, the higher the amount of time required by stimulus $s$ to get a response). In a Linear Model (LM), the expected value of the dependent variable (in this case, the log-time response) of respondent $p$ to stimulus $s$ is directly linked to the linear combination of predictors $\eta_{ps}$ by an *identity* function. The log-normal model (Equation 3 in Table 1) can be equated to a LM (Equation 4 in Table 1). Thus, the estimates of the former model can be obtained from the latter one.

The linear combination of predictors $\eta_{ps}$ of (G)LM can be extended to include fixed and random factors, this leading to the specification of (Generalized) Linear Mixed-Effects Model ((G)LMM). The Best Linear Unbiased Predictors (BLUPs, i.e., the conditional modes of the random effects; [Doran et al., 2007]) describe the deviation of each level of the random effects (i.e., each respondent and each stimulus) from the fixed effects. BLUPs are used for obtaining the Rasch and log-normal model estimates ([De Boeck et al., 2011; Doran et al., 2007; Gelman & Hill] 2007). When Rasch and log-normal model estimates are obtained from (G)LMMs, the relationship between respondent and stimulus parameters changes (i.e., they are summed). Consequently, the difficulty parameter $b_s$ of the Rasch model and the speed parameter $\tau_p$ of the log-normal model are reversely interpreted. The higher the value of $b_s$, the higher the number of correct responses registered on stimulus $s$ (i.e., easiness parameter). The higher the value of $\tau_p$, the larger the amount of time respondent $p$ spends on all items (i.e., slowness parameter).

The Rasch and log-normal parametrizations depend on the random structure of the model, which in turn depends on the variability observed in the data. Different approaches for finding the most appropriate model exist, depending on whether the focus is on the random structure (e.g., [Barr et al., 2013]) or on the fixed one (e.g., [Ryoo, 2011]). The approach followed in this study consists in the comparison of models with increasingly complex random structures (e.g., [Matuschek et al., 2017]).
Fixed and random structures of (G)LMMs

In all models, the fixed intercept is set at 0 (i.e., none of the levels of the fixed slope is taken as the reference category). The fixed and random structures of (G)LMMs are identical. The error variance in the GLMMs follows a logistic distribution \( \varepsilon \sim \text{Logistic}(0, \sigma^2_\varepsilon) \), while that in the LMMs follows a normal distribution \( \varepsilon \sim \mathcal{N}(0, \sigma^2_\varepsilon) \). In what follows, the GLMMs applied to accuracy responses and the LMMs applied to log-time responses are identified by capitals As and Ts, respectively. An overview of the structures of each model is reported in Table 2. Further details are reported in the appendix.

**Table 2**

*Overview of accuracy and log-time models.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed slope</th>
<th>Respondents</th>
<th>Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Implicit measure</td>
<td>Overall ((\theta_p))</td>
<td>Overall ((b_s))</td>
</tr>
<tr>
<td>A2</td>
<td>Implicit measure</td>
<td>Measure-specific ((\theta_{pm}))</td>
<td>Overall ((b_s))</td>
</tr>
<tr>
<td>A3</td>
<td>Associative condition</td>
<td>Condition-specific ((\theta_{pmk}))</td>
<td>Overall ((b_s))</td>
</tr>
<tr>
<td>T1</td>
<td>Implicit measure</td>
<td>Overall ((\tau_p))</td>
<td>Overall ((\delta_s))</td>
</tr>
<tr>
<td>T2</td>
<td>Implicit measure</td>
<td>Measure-specific ((\tau_{pm}))</td>
<td>Overall ((\delta_s))</td>
</tr>
<tr>
<td>T3</td>
<td>Associative condition</td>
<td>Condition-specific ((\tau_{pmk}))</td>
<td>Overall ((\delta_s))</td>
</tr>
</tbody>
</table>

*Note:* Respondent \( p = 1, \ldots, P \), Stimulus \( s = 1, \ldots, S \), Condition \( k = 1, \ldots, K \), Measure \( m = 1, \ldots, M \), where \( P, S, K, \) and \( M \) are the number of respondents, stimuli, conditions, and implicit measures respectively. In all models, the fixed intercept is set at 0.

Considering the intercepts of respondents and stimuli across conditions and implicit measures, Model 1 addresses the between respondents and the between stimuli variabilities across conditions and measures. Model 1 yields overall respondent and stimulus estimates across implicit measures. This model is expected to be the best fitting one when low variability at both levels is observed. This suggests that neither the performance of the respondents nor the functioning of the stimuli change between implicit measures and/or associative conditions.
In Model 2, the random slopes of respondents in implicit measures and the random intercepts of stimuli across measures are specified to address the within–respondents within–measures variability and the between stimuli variability. Model 2 yields measure–specific respondent estimates and overall stimulus estimates. This model is expected to be the best fitting one when a high within–respondents between–conditions variability is observed. This suggests that the performance of the respondents changes according to the implicit measure.

In Model 3, the random slopes of respondents in each associative condition of each implicit measure are specified to account for the within–respondents between–conditions and measures variability. The random intercepts of stimuli across associative conditions and implicit measures are specified as well. This model results in condition–specific respondent estimates for each implicit measure and overall stimulus estimates across implicit measures. Model 3 is expected to be the best fitting model when high within–respondents variability between associative conditions of each implicit measure is observed. This suggests that their performance varies according to the associative condition of each implicit measure. The difference between condition–specific estimates of each implicit measure expresses the bias on the performance of the respondents due to the associative conditions.

**Method**

A Chocolate IAT, a Milk chocolate SC-IAT, and a Dark chocolate SC-IAT were used. Models were fitted in R (R Core Team, 2018) with the lme4 package (Bates et al., 2015) (bobyqa optimizer). The implicitMeasures package (Epifania, Anselmi, & Robusto, 2020b) was used for computing IAT and SC-IAT D scores. Graphical representations were obtained with ggplot2 (Wickham, 2016). Generic R codes for the estimation of accuracy and log-time models are available as supplementary materials.
Data are the same as those in Epifania, Anselmi, and Robusto (2020a), although the aim and the statistical analyses of that study were different from those of the current study. The study in Epifania, Anselmi, and Robusto (2020a) was aimed at a fairer comparison between the IAT and SC-IAT abilities to predict behavioral outcomes by developing new algorithms for the IAT and the SC-IAT in line with typical $D$ scores.

The fit of the data to the model was evaluated with outlier-sensitive fit statistics (i.e., outfit statistics). Outfit statistics are sensitive to unexpected responses observed when the locations of persons and items on the latent trait are far away from each other. These statistics were adjusted to account for both the cross-classified structure of implicit measures and for the specific case of log-response times. Outfit statistics were computed for respondents and stimuli. The procedure for their computation is illustrated in the appendix. Given the peculiar application, ad-hoc thresholds indicating either overfit (i.e., data show less variability than that expected by the model) or underfit (i.e., data show more variability than that expected by the model) were inspired by those in Linacre (2002). Lower and upper bounds of .70 and 1.30 were chosen to indicate overfit and underfit of the data to the model, respectively.

**Participants**

One-hundred and sixty-one people (F = 63.55%, Age = 23.95 ± 2.83 years) were recruited at the University of XXXX. They were informed about the confidentiality of the data and were asked for their consent to take part in the study. Majority of the participants were students (94.08%).

**Materials and procedure**

Twenty-six attributes were used to represent two evaluative dimensions (13 Good and 13 Bad exemplars) and fourteen chocolate images were used to represent two targets (7 Dark and 7 Milk
chocolate exemplars). The same starting seven chocolate images were graphically modified to represent either dark or milk chocolate.

Dark and milk chocolate images were presented in the Chocolate IAT. The critical blocks were composed of 60 trials each, defining the Dark-Good/Milk-Bad condition (DGMB), and the Milk-Good/Dark-Bad condition (MGDB). The SC-IATs employed only either dark (Dark SC-IAT) or milk (Milk SC-IAT) chocolate images. The critical blocks of the SC-IATs were composed of 72 trials each. The critical blocks of the dark SC-IAT were the Dark-Good/Bad (DG) condition and the Good/Dark-Bad (DB) condition. The critical blocks of the milk SC-IAT were the Milk-Good/Bad (MG) condition and the Good/Milk-Bad (MB) condition.

Respondents’ chocolate preferences were explicitly investigated with two items (i.e., “How much do you like milk chocolate?” and “How much do you like dark chocolate?”) evaluated on a 6-point Likert-type scale (from 0 - Not at all to 5 - Very much). At the end of the experiment, participants were offered with a free bar of dark or milk chocolate. The experimenter registered their choices after they left the laboratory. Respondents were tested individually in a laboratory setting. The order of presentation of the implicit measures was counterbalanced across respondents. The explicit evaluation and the choice task were always presented at the end of the experiment.

Data cleaning and D score computation

The IAT was scored with the $D4$ algorithm in Greenwald et al. (2003) (i.e., trials $>10,000$ ms were discarded, incorrect responses were replaced by the average response time inflated by a 600 ms penalty). Positive scores indicate a preference for dark chocolate over milk chocolate. The SC-IAT was scored according to Karpinski and Steinman (2006) (i.e., trials $<350$ ms were discarded, incorrect responses were replaced by the average response time inflated by a 450 ms penalty).
penalty). In both SC-IATs, positive scores indicate a positive evaluation of the target chocolate. The raw response times of each trial were used for estimating the log-normal model (i.e., the error latencies were not assigned any penalty).

Results

Accuracy models

Models A1, A2, and A3 were compared between each other. Model A3 (AIC = 22,365, BIC = 22,618, Log-likelihood = −11,154, Deviance = 22,309) resulted as the best fitting one (Model A1: AIC = 22,991, BIC = 23,036, Log-likelihood = −11,490, Deviance = 22,981, Model A2: AIC = 22,906, BIC = 22,996, Log-likelihood = −11,223, Deviance = 22,886), providing condition–specific respondent ability estimates for each implicit measure (θ_{DGMB}, θ_{MGDB}, θ_{DG}, θ_{DB}, θ_{MG}, and θ_{MB}), and overall stimulus easiness estimates across implicit measures (b_s).

The highest probability of a correct response was observed in MGDB condition (log-odds = 4.05, SE = 0.13, z = 30.92, p < .001), followed by those in DB and that in MG conditions (log-odds = 3.45, SE = 0.11, z = 31.63, p < .001, log-odds = 3.41, SE = 0.10, z = 33.65, p < .001, respectively). The probabilities of a correct response in DG and MB conditions were similar (DG: log-odds = 3.23, SE = 0.10, z = 32.70, p < .001 and MB: log-odds = 3.21, SE = 0.10, z = 31.07, p < .001). The lowest probability of a correct response was observed in DGMB condition (log-odds = 2.92, SE = 0.09, z = 31.89, p < .001). Between–respondents variability was the highest in MGDB (σ^2 = 1.05), followed by those in DG and MB conditions (σ^2_{DG} = 0.84, σ^2_{MB} = 0.75). Respondent variabilities in DB and MG conditions were similar (σ^2_{DB} = 0.63 and σ^2_{MG} = 0.64), while that in DGMB was the lowest (σ^2 = 0.47). Between–stimuli variability was 0.11.

The largest numbers of underfitting respondents were observed in MGDB condition (n =
42) and in MG condition \((n = 27)\). The largest numbers of overfitting respondents were observed in MGDB and MG conditions (MGDB: \(n = 58\) and MG: \(n = 54\)). On average, stimuli showed a good fit to the model (Outfit statistics: \(M = 0.87 \pm 0.14, \text{Min} = 0.70, \text{Max} = 1.17\)).

The distribution of easiness estimates are reported in Figure 1a.

![Figure 1a: Distribution of stimulus easiness](image)

Figure 1: Distribution of stimulus easiness (1a) and time intensity estimates (1b).

Stimuli *Dark1, agony, and annoying* were identified as being excessively difficult in respect to the stimuli of the same category.

**Log-time model**

Models T1, T2, and T3 were compared between each other. Model T3 (AIC = 38,933, BIC = 39,195, Log-likelihood = −19,437, Deviance = 38,875) resulted as the best fitting one (Model T1: AIC = 44,624, BIC = 44,678, Log-likelihood = −22,306, Deviance = 44,612, and Model T2: AIC = 43,448, BIC = 43,548, Log-likelihood = −22,306, Deviance = 44,612), providing condition–specific respondent slowness estimates \(\tau_{DGMB}, \tau_{MGDB}, \tau_{DG}, \tau_{DB}, \tau_{MG}, \tau_{MB}\) and overall stimulus time intensity estimates across implicit measures \(\delta_s\).

Responses in the IAT associative conditions tended to be slower \((B_{DGMB} = −0.12, SE = 0.02, t = −5.63, B_{MGDB} = −0.35, SE = 0.02, t = −19.36\) than in the associative conditions of
both SC-IATs ($B_{DB} = -0.47, SE = 0.02, t = -30.46, B_{DG} = -0.45, SE = 0.02, t = -30.16,$
and $B_{MB} = -0.45, SE = 0.02, t = -29.30$ and $B_{MG} = -0.50, SE = 0.01, t = -34.82$).
Respondents showed higher within and between associative conditions variability in the IAT
($\sigma^2_{DGMB} = 0.05$ and $\sigma^2_{MGDB} = 0.03$), than in both SC-IATs, where their between–conditions
variabilities were identical (Dark SC-IAT: $\sigma^2_{DG} = \sigma^2_{DB} = 0.02$) or almost identical (Milk SC-
IAT: $\sigma^2_{MB} = 0.02$ and $\sigma^2_{MG} = 0.01$). Between–stimuli variability was low ($\sigma^2 = 0.004$).

The highest numbers of underfitting respondents were observed in the two conditions of the
Dark SC-IAT (DB: $n = 29$, DG: $n = 26$). Outfit statistics suggested that all stimuli fitted to
the model ($M = 1.18 \pm 0.03$, $Min= 1.12$, $Max= 1.22$) . The distribution of the time intensity
estimates are reported in Figure [1b] Two stimuli required too much time for getting a response
in respect to the stimuli belonging to the same category. One of these stimuli (i.e., annoying)
resulted to be an extremely difficult stimulus as well.

**Relationship between model estimates, explicit measures, and typical scoring**

Pearson’s correlations between model estimates, explicit measures, and $D$ scores are reported
in Table [3] Explicit chocolate evaluations correlated with typical scoring methods of the IAT
and the Dark SC-IAT, and with condition–specific slowness estimates of these measures. The
direction of these correlations was consistent with the explicitly reported chocolate preference.
The more positive the explicit milk chocolate evaluation, the lower the slowness in MGDB
and DB conditions. The more positive the dark chocolate evaluation, the lower the slowness
in DGMB and DG conditions. No correlations between any of the explicit evaluations and
condition–specific slowness estimates of the Milk SC-IAT or its typical scoring were found.
Grounding on this evidence, the performance at the implicit measures appears to be mostly
influenced by the evaluation of dark chocolate than any evaluation of milk chocolate.
Table 3
Pearson correlations between explicit chocolate evaluations, typical scoring methods, and model estimates.

<table>
<thead>
<tr>
<th></th>
<th>Milk</th>
<th>Dark</th>
<th>D Milk</th>
<th>D Dark</th>
<th>D IAT</th>
<th>( \theta_{DGMB} )</th>
<th>( \theta_{MGDB} )</th>
<th>( \theta_{DB} )</th>
<th>( \theta_{MB} )</th>
<th>( \tau_{DGMB} )</th>
<th>( \tau_{MGDB} )</th>
<th>( \tau_{DB} )</th>
<th>( \tau_{DG} )</th>
<th>( \tau_{MB} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark</td>
<td>-0.39***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D Milk</td>
<td>0.09</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D Dark</td>
<td>-0.20*</td>
<td>0.19*</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D IAT</td>
<td>-0.37***</td>
<td>0.37***</td>
<td>0.01</td>
<td>0.34***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{DGMB} )</td>
<td>-0.15</td>
<td>0.14</td>
<td>0.01</td>
<td>0.16</td>
<td>0.36***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{MGDB} )</td>
<td>0.05</td>
<td>-0.14</td>
<td>0.08</td>
<td>-0.08</td>
<td>-0.23**</td>
<td>0.48***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{DB} )</td>
<td>-0.01</td>
<td>-0.12</td>
<td>-0.05</td>
<td>-0.09</td>
<td>-0.11</td>
<td>0.53***</td>
<td>0.90***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{DG} )</td>
<td>-0.08</td>
<td>0.02</td>
<td>0.12</td>
<td>0.17*</td>
<td>0.15</td>
<td>0.81***</td>
<td>0.68***</td>
<td>0.70***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{MB} )</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.15</td>
<td>-0.1</td>
<td>-0.01</td>
<td>0.54***</td>
<td>0.75***</td>
<td>0.94***</td>
<td>0.63***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_{MG} )</td>
<td>-0.03</td>
<td>-0.08</td>
<td>0.16*</td>
<td>0.12</td>
<td>-0.05</td>
<td>0.58***</td>
<td>0.82***</td>
<td>0.78***</td>
<td>0.91***</td>
<td>0.66***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_{DGMB} )</td>
<td>0.14</td>
<td>-0.31***</td>
<td>0.02</td>
<td>-0.19*</td>
<td>-0.59***</td>
<td>-0.20*</td>
<td>0.23***</td>
<td>0.14</td>
<td>-0.02</td>
<td>0.09</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_{MGDB} )</td>
<td>-0.27***</td>
<td>0.11</td>
<td>0.04</td>
<td>0.15</td>
<td>0.36***</td>
<td>0.23**</td>
<td>0.10</td>
<td>0.14</td>
<td>0.19*</td>
<td>0.17*</td>
<td>0.13</td>
<td>0.42***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_{DB} )</td>
<td>-0.20*</td>
<td>0.02</td>
<td>0.07</td>
<td>0.33***</td>
<td>0.24**</td>
<td>0.22**</td>
<td>0.15</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
<td>0.12</td>
<td>0.37***</td>
<td>0.77***</td>
<td></td>
</tr>
<tr>
<td>( \tau_{DG} )</td>
<td>-0.05</td>
<td>-0.16*</td>
<td>0.02</td>
<td>-0.17*</td>
<td>-0.01</td>
<td>0.12</td>
<td>0.18*</td>
<td>0.15</td>
<td>0.09</td>
<td>0.14</td>
<td>0.08</td>
<td>0.61***</td>
<td>0.73***</td>
<td>0.80***</td>
</tr>
<tr>
<td>( \tau_{MB} )</td>
<td>-0.06</td>
<td>-0.08</td>
<td>0.26**</td>
<td>0.01</td>
<td>0.13</td>
<td>0.21*</td>
<td>0.24**</td>
<td>0.24**</td>
<td>0.19*</td>
<td>0.22*</td>
<td>0.17*</td>
<td>0.43***</td>
<td>0.72***</td>
<td>0.77***</td>
</tr>
<tr>
<td>( \tau_{MG} )</td>
<td>-0.10</td>
<td>-0.10</td>
<td>-0.21**</td>
<td>-0.04</td>
<td>0.09</td>
<td>0.16*</td>
<td>0.17*</td>
<td>0.19*</td>
<td>0.15</td>
<td>0.20*</td>
<td>0.12</td>
<td>0.48***</td>
<td>0.72***</td>
<td>0.72***</td>
</tr>
</tbody>
</table>

Note: ***: \( p < .001 \), **: \( p < .01 \), *: \( p < .05 \). \( \theta \): Ability estimates, \( \tau \): Sloveness estimates, DGMB: Dark/Good-Milk/Bad condition (IAT), MGDB: Milk/Good-Dark/Bad condition (IAT), DG: Dark/Good condition (Dark SC-IAT), DB: Dark/Bad condition (Dark SC-IAT), MG: Milk/Good condition (Milk SC-IAT), MB: Milk/Bad condition (Milk SC-IAT).
The higher correlations between one of the condition–specific slowness estimates of each implicit measure and their respective typical scoring suggested that the associations in only one of the associative conditions played a major role in influencing the performance (IAT: \( z = 12.82, p < .001 \), Dark SC-IAT: \( z = 10.60, p < .001 \) and Milk SC-IAT: \( z = 10.43, p < .001 \)). The correlations between the typical scoring methods and ability estimates indicated a substantial contribution of ability in performing at the categorization task.

In the IAT case, the higher the ability in DGMB condition, the slower the respondents in the same condition and the faster in the opposite condition. Moreover, respondents with higher ability in MGDB condition tended to have a higher slowness in DGMB condition. Ability in MGDB condition was not correlated with the slowness in the same condition.

**Prediction of a behavioral outcome**

The predictive abilities of Rasch and log-normal model estimates and of typical scoring methods were compared. Differential measures (i.e., typical scoring methods, time differentials – difference between condition–specific slowness estimates– and ability differentials –difference between condition–specific ability estimates–), and linear combinations of their single components (i.e., average response time in each associative condition of each measure –D scores– or condition–specific ability and slowness estimates –Rasch and log-normal model estimates–) were considered. Ability and time differentials can be considered as accuracy-based and latency-based measures of the IAT effect, respectively.

Four models were specified including the predictors of interest (i.e., typical scoring methods vs. model estimates, differential measures vs. linear combinations of their single components). Relevant predictors were chosen with forward selection. *Nagelkerke’s R*² ([Nagelkerke, 1991]) was computed as *Pseudo R*². To investigate the linear combinations of predictors that best
accounted for the choice, model general accuracy of prediction (i.e., proportion of choices correctly identified by the model), dark chocolate choices (DCCs) accuracy (i.e., proportion of DCCs correctly identified by the model), and milk chocolate choices (MCCs) accuracy (i.e., proportion of MCCs correctly identified by the model) were computed. Results of stepwise logistic regressions are reported in Table 4. All models explained about the same proportion of variance, ranging from 0.12 (Model 2) to 0.19 (Model 4). Only the IAT $D$ score, its single components, and its time–differentials (Models 1, 2, and 3) were deemed relevant for the prediction. Slowness estimates of both IAT conditions and of DG condition (Dark SC-IAT) were identified as relevant predictors (Model 4).

<table>
<thead>
<tr>
<th>Model</th>
<th>log-odds</th>
<th>SE</th>
<th>Nagelkerke $R^2$</th>
<th>Gen</th>
<th>DCC</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>Intercept</td>
<td>$-0.08$</td>
<td>$0.16$</td>
<td>$0.00$</td>
<td>$0.48$</td>
<td>$0.00$</td>
</tr>
<tr>
<td></td>
<td>Typical scoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Intercept</td>
<td>$-1.03^{***}$</td>
<td>$0.31$</td>
<td>$0.16$</td>
<td>$0.64$</td>
<td>$0.62$</td>
</tr>
<tr>
<td></td>
<td>$D$ score</td>
<td>$-1.55^{***}$</td>
<td>$0.39$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Intercept</td>
<td>$-0.36$</td>
<td>$0.91$</td>
<td>$0.12$</td>
<td>$0.62$</td>
<td>$0.67$</td>
</tr>
<tr>
<td></td>
<td>$M_{DGMB}$</td>
<td>$0.01^{**}$</td>
<td>$0.01$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M_{MGDB}$</td>
<td>$-0.01^{*}$</td>
<td>$0.01$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Model estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Intercept</td>
<td>$-0.95^{***}$</td>
<td>$0.29$</td>
<td>$0.15$</td>
<td>$0.64$</td>
<td>$0.67$</td>
</tr>
<tr>
<td></td>
<td>($\tau_{MGDB} - \tau_{DGMB}$)</td>
<td>$-3.57^{***}$</td>
<td>$0.91$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Intercept</td>
<td>$0.44$</td>
<td>$0.81$</td>
<td>$0.19$</td>
<td>$0.64$</td>
<td>$0.65$</td>
</tr>
<tr>
<td></td>
<td>$\tau_{DGMB}$</td>
<td>$2.45^{*}$</td>
<td>$1.08$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_{MGDB}$</td>
<td>$-5.99^{***}$</td>
<td>$1.76$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau_{DG}$</td>
<td>$5.29^{*}$</td>
<td>$2.59$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$. $\theta$: Ability estimates, $\tau$: Slowness estimates, DGMB: Dark/Good-Milk/Bad condition (IAT), MGDB: Milk/Good-Dark/Bad condition (IAT), DG: Dark/Good condition (Dark SC-IAT), DB: Dark/Bad condition (Dark SC-IAT), MG: Milk/-Good condition (Milk SC-IAT), MB: Milk/Bad condition (Milk SC-IAT), $Gen$: General accuracy, $DCC$: Dark chocolate choice accuracy, $MCC$: Milk chocolate choice accuracy.
Final remarks

The study was aimed at introducing a modeling framework for a Rasch analysis of multiple implicit measures administered concurrently. The focus was on the specific case of the IAT administered together with the SC-IAT. The results suggest Linear Mixed-Effects Models (LMMs) allow for obtaining a Rasch parametrization of IAT and SC-IAT accuracy and time responses while accounting for the within and between measures sources of variability. The detailed information at stimulus and respondent levels allows for delving deeper on the meaning and functioning of implicit measures, as well as for grasping a better understanding of the processes underlying behavioral decisions.

The pattern of correlations between condition–specific ability and slowness estimates of the IAT suggests a speed-accuracy trade-off, according to which respondents who performed better (i.e., higher ability) in categorizing the stimuli in DGMB condition needed to slow down in the opposite condition to prevent a loss in the accuracy performance. In the SC-IAT cases, slowness and ability estimates were either not correlated (Dark SC-IAT) or positively correlated (Milk SC-IAT). Consistent with previous findings (see, e.g., Karpinski & Steinman, 2006), this suggests that the SC-IAT is an easier task than the IAT.

When differential measures were used for predicting the choice, only the information provided from the IAT was deemed relevant. Consistently, only condition–specific average response times of the IAT were relevant predictors for the choice. However, when the linear combination of condition–specific estimates were considered, also the slowness in Dark-Good condition of the Dark SC-IAT was deemed relevant for the choice prediction. Although the model including the speed in Dark-Good condition did not result in a higher predictive accuracy of the choice or a higher proportion of explained variance, it provided a better understanding of the processes underlying the choice of the respondents. This result allowed for speculating
that the choice was mostly driven by the positive evaluation (i.e., preference) for dark chocolate than any association with milk chocolate. The contribution of the Dark SC-IAT was lost when typical scoring methods, differential measures, and single components of typical scoring methods were used for the prediction. This modeling framework might be of particular use for investigating the processes underlying inter-group behaviors. For instance, it might help in understating whether the decision to affiliate with members of stigmatized groups is mostly influenced by in-group favoritism or out-group derogation.

The information at the stimulus level is noteworthy as well. Stimulus outliers for each stimulus categories were highlighted. This suggests a malfunctioning that could bias systematically the performance of the respondents, such that the results might not reflect the automatic associations of the respondents but the bias introduced by some of the stimuli. The stimuli flagged as malfunctioning by easiness and time intensity estimates do present characteristics that might undermine their correct and fast categorization, which are mostly related to the Italian lexicon they belong. De Mauro (2016) decomposed the basic Italian lexicon (i.e., the lexicon used for every-day spoken and written interactions) into three categories, namely the high availability lexicon (i.e., lexicon used only in specific contexts but accessible by the largest part of the population), the fundamental lexicon (i.e., the most used lexicon with which people are most familiar), and the high-frequency lexicon (i.e., the lexicon learned in school, easily accessible although less used than the fundamental one). Agony (English translation of agonia) and noxious (English translation of nocivo) belongs to the high availability lexicon, but only agony showed a malfunctioning. Indeed, noxious is usually employed to indicate an health hazard associated with food, making its negative meaning more immediately accessible. On the other hand, the negative valence of agony might be less immediately accessible, both because of its use in specific contexts (e.g., medical) and/or because it might more immediately recall the
idea of pain and suffering. Most of other attribute stimuli belonged to either the fundamental lexicon or the high-frequency lexicon. Although annoying (i.e., English translation of the Italian fastidioso) belongs to the high-frequency lexicon, it showed a time intensity malfunctioning (i.e., too much time in respect to its own category to get a response). Trivially, its higher time intensity estimate might be due to the length of the word itself.

The outfit statistics presented in this work are a first attempt at providing an absolute fit statistic to these models. They potentially allow for pinpointing stimuli that are misinterpreted by the respondents or respondents that did not perform according to the instructions. The potential of the outfit statistics introduced in this study should be further investigated in future studies, as well as their validity as absolute fit statistics.

The application of the modeling framework introduced in this contribution was focused on the specific case of the IAT administered with the SC-IAT. As discussed above, it is not usual to find studies where multiples implicit measures are administered together (e.g., Green et al., 2019; Miles et al., 2019). Given the flexibility of Linear Mixed-Effects Models (LMMs), this modeling framework can be easily adapted to these instances.

In conclusion, this work supported the feasibility of LMMs for obtaining useful information at both respondent and stimulus levels in a Rasch framework. Additionally, it represents a first step towards a comprehensive modeling of implicit measures administered together.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest
Data availability statement

Data are available in the public Open Science Framework repository [https://osf.io/7qm8p/](https://osf.io/7qm8p/)
References


10.1111/j.1467-9868.2007.00600.x


10.4473/TPM27.2.4


10.31124/advance.11914416.v1


10.31124/advance.13227713.v1


error and power in linear mixed models. *Journal of Memory and Language, 94*, 305-315. doi: https://doi.org/10.1016/j.jml.2017.01.001


Appendix

Models specification

Respondent and stimulus estimates of Rasch and log-normal models are obtained from Best Linear Unbiased Predictors (BLUPs) of (Generalized) Linear Mixed-Effects Models (G)LMMs. BLUPs describe the deviation of each level of the random factor from the estimates of the fixed factor [De Boeck et al., 2011; Doran et al., 2007]. The person parameters ($\theta_p$ or $\tau_p$) are derived from the random effects of respondents. In Model 1, the random effect of respondents is uni-dimensional ($\alpha_p \sim \mathcal{N}(0, \sigma^2_{\alpha_p})$). BLUPs describe the deviation of respondents from the fixed intercept. In Model 2, the random effect of respondents is extended to be three-dimensional (i.e., one dimension for each implicit measure) with a general variance-covariance matrix ($\beta_{pm} \sim \mathcal{MVN}(0, \Sigma_{pm})$). BLUPs define the deviation of each respondent from the fixed estimate of each implicit measure. In Model 3, the random effect of respondents is allowed to be six-dimensional (i.e., one dimension for each condition of each implicit measure) with a general variance-covariance matrix ($\beta_{pm} \sim \mathcal{MVN}(0, \Sigma_{pk})$). BLUPs define the deviation of each respondent from the fixed estimate of the associative condition of each implicit measure. The stimulus parameters ($b$ or $\delta$) derive from the random effects of stimuli. In all models, the random effects of stimuli are uni-dimensional ($\alpha_s \sim \mathcal{N}(0, \sigma^2_{\alpha_s})$), and BLUPs describe their deviation from the fixed intercept.

Accuracy models

The inverse of the logit link function ($logit = \frac{\mu_{ps}}{1-\mu_{ps}}$, where $\mu_{ps}$ is the probability of a correct response by respondent $p$ to stimulus $s$) is $logit^{-1}$. This yields the probability of a correct response given the linear combination of predictors $\eta_{ps}$, which include both fixed and random effects ($\mu_{ps} = logit^{-1} = \frac{\exp(\eta_{ps})}{1+\exp(\eta_{ps})}$). The error variance follows a logistic distribution (i.e., $\varepsilon \sim \text{Logistic}(0, \sigma^2)$).

Model A1: The fixed slope is the implicit measure $m$. The random intercepts of respondents ($p = 1, \ldots, P$) and stimuli ($s = 1, \ldots, S$) across associative conditions ($k = 1, \ldots, K$) and
implicit measures \((m = 1, \ldots, M)\) are specified:

\[
y_i = \logit^{-1}(\alpha + \beta_m X_m + \alpha_p[i] + \alpha_s[i] + \epsilon_i),
\]

with \(\alpha_p \sim \mathcal{N}(0, \sigma_p^2)\) and \(\alpha_s \sim \mathcal{N}(0, \sigma_s^2)\). The random structure of Model A1 yields overall respondent ability \(\theta_p\) and overall stimulus easiness \(b_s\) estimates.

**Model A2:** The random slopes of respondents in implicit measures and the random intercepts of stimuli are specified:

\[
y_i = \logit^{-1}(\alpha + \beta_m X_m + \alpha_k[i] + \beta_p[i] m_s + \epsilon_i),
\]

with \(\beta_{pm} \sim \mathcal{MN}(0, \Sigma_{pm})\) (where \(\Sigma_{pm}\) is the variance-covariance matrix of the population of the respondents) and \(\alpha_s \sim \mathcal{N}(0, \sigma_s^2)\). Model A2 yields measure–specific respondent ability estimates \(\theta_{pm}\) and overall stimulus easiness \(b_s\) estimates.

**Model A3:** The associative conditions of each implicit measure are specified as the fixed slope. The random slopes of respondents in associative conditions of each implicit measure and the random intercepts of stimuli across conditions and implicit measures are specified:

\[
y_i = \logit^{-1}(\alpha + \beta_k X_k + \alpha_s[i] + \beta_p[i] k_s + \epsilon_i),
\]

with \(\alpha_s \sim \mathcal{N}(0, \sigma_s^2)\) and \(\beta_{pk} \sim \mathcal{MN}(0, \Sigma_{pk})\) (where \(\Sigma_{pk}\) is the variance-covariance matrix of the population of respondents). Model A3 yields condition and measure specific ability estimates \(\theta_{pmk}\) and overall stimulus easiness \(b_s\) estimates.

**Log-time models specification**

The linear combination of predictors \(\eta_{ps}\) is directly linked to the observed log-time response through the *identity* function. The error variance follows a normal distribution (i.e., \(\epsilon \sim \mathcal{N}(0, \sigma^2)\)). Overall stimulus time intensity estimates \(\delta_s\) are obtained from all models.

**Model T1:** The fixed slope is the implicit measure \(m\). The random intercepts of respondents \(p = 1, \ldots, P\) and stimuli \(s = 1, \ldots, S\) across associative conditions \(k = 1, \ldots, K\) and
implicit measures $m = 1, \ldots, M$ are specified:

$$y_i = \alpha + \beta_m X_m + \alpha_p[i] + \alpha_s[i] + \epsilon_i, \quad (8)$$

with $\alpha_p \sim \mathcal{N}(0, \sigma^2_p)$ and $\alpha_s \sim \mathcal{N}(0, \sigma^2_s)$. The random structure of Model T1 yields overall respondent slowness $\tau_p$ and overall stimulus easiness $b_s$ estimates.

**Model T2:** The random slopes of respondents in implicit measures and the random intercepts of stimuli are specified:

$$y_i = \alpha + \beta_m X_m + \alpha_k[i] + \beta_p[i] m_i + \epsilon_i, \quad (9)$$

with $\beta_{pm} \sim \mathcal{MVN}(0, \Sigma_{pm})$ (where $\Sigma_{pm}$ is the variance-covariance matrix of the population of respondents) and $\alpha_s \sim \mathcal{N}(0, \alpha^2_s)$. Model T2 yields measure–specific respondent slowness $\tau_{pm}$ and overall stimulus easiness $b_s$ estimates.

**Model T3:** The associative conditions of each implicit measure are specified as the fixed slope. The random slopes of respondents in the associative conditions of each implicit measure and the random intercepts of stimuli across conditions and implicit measures are specified:

$$y_i = \alpha + \beta_k X_k + \alpha_s[i] + \beta_{pk[i]} k_i + \epsilon_i, \quad (10)$$

with $\alpha_s \sim \mathcal{N}(0, \alpha^2_s)$ and $\beta_{pk} \sim \mathcal{MVN}(0, \Sigma_{pk})$ (where $\Sigma_{pk}$ is the variance-covariance matrix of the population of the respondents). Model T3 yields condition and measure specific respondent slowness $\tau_{pkm}$ and overall time intensity $\delta_s$ estimates.

**Outlier-sensitive fit statistics**

**Standardized residuals of the Rasch model**

The standardized residuals of the Rasch model are usually computed as

$$z = -\sqrt{\exp(\theta_p - b_s)} \quad (11)$$
when \( x_{ps} = 0 \) and

\[
z = \sqrt{e^{xp(b_s - \theta_p)}}
\]  
(12)

when \( x_{ps} = 1 \).

When GLMMs are used to estimate the Rasch model, respondent and stimulus parameters are summed. Thus, the standardized residuals must be computed as follows:

\[
z = -\sqrt{e^{xp(-b_s - \theta_p)}}
\]  
(13)

when \( x_{ps} = 0 \) and

\[
z = \sqrt{e^{xp(\theta_p + b_s)}}
\]  
(14)

when \( x_{ps} = 1 \).

**Standardized residuals of the log-normal model**

To the best of our knowledge, no procedures are available for computing outfit statistics of the log-normal model, hence we adapted the procedure used for the Rasch model.

The difference between respondent and stimulus parameters \( \tau_p \) and \( \delta_s \) is computed according the comparison between the observed log-response to each trial \( t_{ps} \) and the value predicted by the model \( \hat{t}_{ps} \):

\[
z = -\sqrt{e^{xp(-\tau_p - \delta_s)}}
\]  
(15)

when \( t_{ps} \leq \hat{t}_{ps} \).

\[
z = \sqrt{e^{xp(\delta_s + \tau_p)}}
\]  
(16)

when \( t_{ps} > \hat{t}_{ps} \).
Outfit computation

Outfit statistics result from the mean of the squared standardized residuals either across subjects (stimuli outfit) or across stimuli (respondents outfit). In this application, the mean is computed according to the best fitting random structure. For instance, if the best fitting model yields condition–specific respondent estimates and overall stimulus estimates, the (condition–specific) respondent outfit statistics are obtained by averaging the squared standardized residuals across stimuli in each condition of each implicit measure. The (overall) stimulus outfit statistics are obtained by averaging the squared standardized residuals across respondents, conditions, and implicit measures.