Assessing Personality Traits Through Response Latencies Using Item Response Theory

Jochen Ranger[1] and Tuulia M. Ortner[2]

Abstract

Recent studies have revealed a relation between the given response and the response latency for personality questionnaire items in the form of an inverted-U effect, which has been interpreted in light of schema-driven behavior. In general, more probable responses are given faster. In the present study, the relationship between the probability of the given response and the response latency was investigated. First, a probabilistic model was introduced describing the relationship between response latencies and a latent trait. Second, the model was applied in an empirical study: Employing items from a personality questionnaire and using data from 170 men, the probability of responses were estimated based on the Rasch model. Assuming log-normally distributed response latencies, a linear regression model was fit to the logarithmized response latencies, including the response probability as a predictor. Findings suggested that the quantities are negatively related. This relation can be used to incorporate the response latency into the estimation of trait levels. For the scales used in the study, the results showed that test information could be increased by 13% to 17% when considering response latencies.

Response latencies have attracted interest in personality psychology as well as in psychological testing right from the beginning: Cattell's (1886) early experiments already dealt with the time needed to name objects, a task that was later adopted to develop the so-called Stroop task (Stroop, 1935). This contribution to the assessment of

[1]Giessen University, Giessen, Germany
[2]Free University Berlin, Berlin, Germany

Corresponding Author:
Jochen Ranger, Giessen University, Otto-Behaghel-Strasse 10, Giessen, Germany
Email: jochen.m.ranger@psychol.uni-giessen.de

personality was demonstrated by showing that response latencies increased in anxious persons when threat words versus neutral words interfered with a color-naming task (Mathews & MacLeod, 1985). Furthermore, in terms of the interpretation of projective tests, for example, in the Rorschach inkblot test, response latencies have been interpreted as an indicator for test takers' emotional arousal (Höhn, 1959). Recently, response latencies were rediscovered and have experienced a return in psychological research by a set of methods using them as indicators for individual differences in the strength of association between cognitive representations. Several tests have been devised for the assessment of implicit attitudes and dispositions, the best known of these methods being the Implicit Association Tests (Greenwald, McGhee, & Schwartz, 1998).

Through the broader use of computers for item presentation, response latencies are usually recorded by default and can be assessed economically. However, the use of response latencies in questionnaires has not yet found its way into practical application in psychological assessment, and the use of questionnaires is still the most common approach for the assessment of personality. In the present study, we introduce a new methodological approach for modeling response latencies based on item response theory (IRT), and we demonstrate its utility for the purpose of psychological assessment in an empirical data application.

Response Latencies in Questionnaires

Response latencies in questionnaires started to arouse interest in the 1970s. At that time, studies mostly dealt with psychometric properties of items depending on average response latencies (Stricker & Alderton, 1999). Research revealed that items with long latencies were ''unstable''—they evoked different responses over time. This has been shown for different questionnaires and for different testing intervals (Dunn, Lushene, & O'Neil, 1972; Holden & Fekken, 1990; Holden, Fekken, & Jackson, 1985). This result was discussed as a function of item difficulty in that items causing long item response latencies were supposed to be ambiguous and less readable. Subsequent studies have confirmed that response latencies provide information about item characteristics, such as item length (Dunn et al., 1972; Neubauer & Malle, 1997). Research has revealed also that average response latencies depend on an item's emotional evocative character (Temple & Geisinger, 1990; Tyron & Mulloy, 1993). Anastasi (1976) additionally highlighted the effect of the testing condition on test behavior and described certain preconditions for a valid measurement as a quiet room free from distractions with adequate temperature, light, and air conditioning. These are all factors that might influence response times in personality questionnaires.

Later, two additional categories were established for information included in response latencies. First, response latencies were shown to reflect certain test-taker characteristics independent of the construct assessed by the items; for example, the individual disposition to answer honestly or to fake (Holden, 1995; Holden & Hibbs, 1995; Holden & Kroner, 1992). Second, recent approaches have addressed the value of response latencies as an additional representation of the personality trait that is

intended to be assessed by the items. Holden, Fekken, and Cotton (1991) investigated response latencies of students and psychiatric patients in the Basic Personality Inventory (BPI; Jackson, 1989). They found that mean decision times for endorsed items were negatively correlated with relevant self-report scale scores and criterion ratings by significant others, whereas mean response latencies for rejected items were correlated positively with both corresponding self-report scale scores and external criteria. In another study addressing response latencies in the assessment of psychopathology traits as well as ''normal personality traits,'' Fekken and Holden (1992) found similar results and interpreted them as support for the hypothesis that response latencies indicate the existence of a person's network of self-knowledge, seeing the self as a cognitive prototype (Kuiper, 1981). Hence, the representation of a person's personality traits was seen as a cognitive network, and results on response latencies were in line with findings that the existence of self-schemata facilitates the processing of information about the self (Markus, 1977). Persons represented at either end of a trait dimension of a personality construct have been shown to respond faster to corresponding items; whereas on the other hand, items representing extreme trait values have revealed shorter response latencies than items in the medium range (Amelang, Eisenhut, & Rindermann, 1991). This phenomenon has been described as the inverted-U effect (Kuiper, 1981), indicating faster response latencies when adjectives were rated as extremely unlike or like the self compared with the ratings of adjectives that were only moderately prototypical.

Studies followed that aimed to further examine the values of response latencies for psychological assessment: Holden and Fekken (1993) investigated psychometric properties of response latencies of 92 university students for 500 items using the Personality Research Form (PRF; Jackson, 1994) measuring five personality dimensions. They found appropriate levels of internal consistency and reliability as well as convergent (*discriminant*) validity of response latencies with self-and peer reports on corresponding (noncorresponding) personality traits. Based on these findings, they postulated that response latencies possess construct validity. More recently, Akrami, Hedlund, and Ekehammar (2007) investigated the relationship between participants' responses to personality items and their response latencies using the NEO-PI Big Five personality inventory (Costa & McCrae, 1985). Akrami again found an inverted-U effect: Persons with a high or a low score on a scale responded faster to personality items assigned to these scales, whereas persons with a score around the mean showed longer response latencies.

In sum, the empirical findings have corroborated the idea that response latencies are related to a person's individual trait level. However, the more or less descriptive database of the preceding studies does not provide possibilities for the practical use of response latencies for personality assessment. For practical implementation, a measurement model is needed to incorporate response latencies into the scoring of the test. Or to be more precise, a formal psychometric model is required to enable the utilization of response latencies on an individual level to estimate a person's trait level. Such models have been developed in the field of achievement tests. For example, van der Linden (2008) revealed the possibility of integrating response latencies into the estimation of

ability levels of individuals. However, in the field of personality testing, the current state of research is unsatisfactory. Approaches from achievement tests cannot be generalized to personality tests because the relation between response latencies and the underlying trait level is different: In achievement tests, one usually observes a monotonic relation between the average response latency and the test score (Lavergne & Vigneau, 1997; MacLennan, Jackson, & Bellantino, 1988; Rafaeli & Tractinsky, 1991) and not the inverted-U effect of personality tests. Hence, different models are needed.

Although a first model for response latencies in personality test items has been proposed by Ferrando and Lorenzo-Sevas (2007), the application of this model might not be completely adequate with reference to some of the presented findings. By using the item–person distance on the latent continuum as the only content-specific determinant of the response latency, this model predicts the same response latency regardless of the given response (endorsement or rejection). For example, according to the model of Ferrando and Lorenzo-Sevas (2007), an individual with a high trait level should generally respond quickly to items located at the lower pole of the trait continuum, even though the response is an improbable rejection. This recent model of Ferrando and Lorenzo-Sevas (2007) is therefore not able to explain or support the different relations between the test score and the average response latency in endorsed and rejected items found by Fekken and Holden (1992) and Holden et al. (1991).

Therefore, in the following, we reinterpret previous findings (Austin, 2009; Holden et al., 1991) in light of modern item response theory (IRT) (Embretson & Reise, 2000; Lord, 1980). First, we introduce a new probabilistic model to describe the contribution of response latencies in personality questionnaire items for trait estimation; this model provides the opportunity to make this information useful for the application in individual assessment. In the second step, we apply this model using data from a questionnaire and assess model fit. In the third step, the benefit of model application with reference to the increase of gained information of the test will be discussed.

An IRT Model for Response Latencies in Personality Questionnaires

Response latencies in test items can be modeled with the log-normal distribution (Schnipke & Scrams, 2002; Thissen, 1983; van der Linden, 2006). This distribution is skewed and defined for positive values only such that it shares important features with empirical response latency distributions. Nevertheless, the log-normal distribution can easily be transformed into the standard normal distribution by calculating the logarithm. This means that in the case of log-normally distributed response latencies, the log response latencies are normally distributed with mean μ and variance σ^2. When combining regression models with the log-normal distribution, the standard approach consists of modeling the expectation of the log-transformed observations as a function of certain predictors while assuming the variance σ^2 to be constant. This amounts to an accelerated failure time model that enjoys great popularity in the field of survival analysis (Klein & Moeschberger, 1997). The application of this approach

to model response latencies in test items is straightforward and consists of tracing back the expected log response latencies to characteristics of the item and of the individual. In the following section, three predictors will be considered.

Based on findings of Holden and colleagues (Fekken & Holden, 1992; Holden & Kroner, 1992; Popham & Holden, 1990), individuals with a high scale score (and hence a high trait level) can endorse items faster than individuals with a low scale score (and hence a low trait level). Individuals with a high scale score (and hence a high trait level) have a high probability of endorsement, whereas individuals with a low scale score (and hence a low trait level) have a low probability of endorsement. Therefore, one can suspect that the probability of the given response might be related to the response latency such that probable responses are faster than less probable responses. The same implication can be derived from rejected items. Individuals with a low scale score (and hence a low trait level) reject items faster than individuals with a high scale score (and hence a high trait level). Individuals with a low trait level, however, have a high probability of rejecting an item, whereas individuals with a high trait level have only a low probability of rejection. Again, one can conclude that it is the probability of the response that is related to the speed of responding such that probable responses are given faster than improbable ones. Therefore, when the test is scaled with an IRT model such that the response probability can be estimated, one would expect a relation between the response latency and the response probability. Since the probability of the response depends on the trait level of an individual, this assumption connects the response latency to the trait with the response probability as the mediating conjunction.

In addition to the probability of a response, further influencing factors on response latencies have to be considered. Before the statistical analysis of response latencies, it is standard practice to center them within each subject and each item (see, e.g., Fekken & Holden, 1992, for a detailed description of this procedure). This adjustment is driven by the intention to remove a large first factor reflecting individual differences in work pace (depending on reading speed) and the hope of reducing the perturbing influence of item characteristics such as length, wording, or ambiguity, such that the relation of the response latencies to the target trait becomes more evident. However, instead of implicitly removing these perturbing sources of influences, they can be accounted for explicitly by introducing them as two additional determinants of an individual's item response latency, namely, the general response speed ω of an individual and the time demand of the item β_0.

Let a scale consisting of binary items be devised to measure the target trait θ, such that the probability of endorsement or rejection of the scale items depends on θ. When implementing the considerations delineated above in a formal regression model, the expected log response latency of an individual in an item should follow the relation

$$E[\log(t)|\theta, \omega, x] = \beta_0 + \omega + \beta_1 P(x|\theta), \qquad (1)$$

where θ is the trait level of the individual, ω is his general response speed, $P(x|\theta)$ is the probability of the given response x, β_0 is the time demand of the item, and β_1 is an item-specific regression weight. Of course, the probability $P(x|\theta)$ is not known.

However, it can be modeled by any IRT model. As the two-parameter logistic model of Birnbaum (1968) can be considered to be a standard model for tests, it will be used in the following.

According to the two-parameter logistic model, the probability of a positive response can be traced back to the trait level θ via the item characteristic curve:

$$P(x = 1|\theta) = \frac{\exp[\alpha_0 + \alpha_1 \theta]}{1 + \exp[\alpha_0 + \alpha_1 \theta]}. \tag{2}$$

In Equation (2), parameter α_0 represents the item difficulty and the parameter α_1 is the item discrimination. In latent trait models, a common assumption claims that observable variables are independent when conditioning on the latent traits. This is justified by the argument that the latent traits are the only systematic factors of influence and should, therefore, account for all relations between the observable variables. Applying this assumption to the present context, one can conclude that after conditioning on θ and ω, the joint distribution of the response and the response latency of an item can be stated as

$$f(\log(t), x|\theta, \omega) = f(\log(t)|\theta, \omega, x) f(x|\theta), \tag{3}$$

where $f(x|\theta)$ is a Bernoulli distribution with success probability $P(x = 1|\theta)$ according to Equation 2 and $f(\log(t)|\theta, \omega, x)$ represents a normal distribution with expectation $E[\log(t)|\theta, \omega, x]$ according to Equation 1 and variance σ^2. The joint distribution of all responses and response times for a test can be derived from the conditional independence assumption in the same way.

Equation (3) or its generalization to the different items of the test is a formal model that relates the responses and the response latencies to the target trait θ. It can be used to infer the individual trait level after observing the responses and the response latencies of an individual. One procedure for calculation is maximum likelihood estimation. This amounts to choosing the trait levels θ and ω with the highest plausibility given the observed data. As traits are usually estimated by responses alone, one might pose the question about the benefits of additionally considering response latencies. This question will be discussed in the following section.

Information of Response Latencies

A central quantity in test theory is the information of the test. The importance of the test information can be explained among other things by the fact that the variance of the maximum likelihood estimator of a trait is inversely related to the test information. Therefore, the utility of the response latencies will be judged by the extent to which the test information can be increased by taking the response latencies into consideration.

In the preceding section, it was assumed that a scale with n binary items can be modeled with the two-parameter logistic model (see Equation 2). All item parameters are known or rather have been estimated precisely with a calibration sample. After

observing the response pattern of an individual, it is possible to estimate the corresponding trait level with maximum likelihood estimation. This amounts to determining the value $\hat{\theta}$ of the trait that maximizes the likelihood of the trait level given the observed response pattern of the individual. For scales of appropriate length, this estimate is approximately normally distributed. The asymptotic variance of the estimator is the reciprocal of the information of the test. When observations are independent, the information of the test is simply the sum of the information of the item responses. For the two-parameter logistic model, the information of an item response is as follows:

$$I_x(\theta) = \alpha_1^2 \frac{\exp[\alpha_0 + \alpha_1\theta]}{[1 + \exp[\alpha_0 + \alpha_1\theta]]^2}. \tag{4}$$

The information of an item response is a function of the trait θ, which is peaked and decreases with increasing/decreasing trait levels (see Baker & Seock-Ho, 2004, for more details). The information of the test can be derived by summing all item information curves of the single items.

Relating the response latencies to the trait level of an individual via Equation (1) enables one to base an estimate of the trait level on both responses and response latencies. Again, trait levels can be estimated by maximum likelihood estimation, exploiting Equation (3) to formulate the likelihood equations. However, two complications have to be accounted for. First, within a single item, the response latency is not independent of the response because the distribution of the response latency depends on $P(x|\theta)$. Second, response latencies depend on two individual characteristics, namely, the trait level of an individual θ and his or her general response speed ω, such that two individual characteristics have to be estimated.

For dependent observations, it is well known that the information of all observations can be decomposed into the information of the conditional distributions (see Sorensen & Gianola, 2002, p. 129, for more details). In the present context, this means that the information of an item consists of the information of the item response, that is, $I_x(\theta)$ of Equation 4 plus the information of the response latency distribution $f(\log(t)|x; \theta, \omega)$ when conditioning on the response. As two traits have to be estimated, the target trait θ and the equally unknown response speed ω, the information of the response latency distribution is a two-dimensional matrix $I_t(\theta, \omega)$. The first entry of the matrix $I_{t[11]}(\theta, \omega)$ could be denoted as the information of a single response latency about the trait level θ. With the assumption of log-normally distributed response latencies, this element can be expressed as

$$I_{t[11]}(\theta, \omega) = \frac{\alpha_1^2 \beta_1^2}{\sigma^2} \left[\frac{\exp[\alpha_0 + \alpha_1\theta]}{[1 + \exp[\alpha_0 + \alpha_1\theta]]^2} \right]^2. \tag{5}$$

Again, the information of the test can be derived by summing the information functions of all n items, as responses and response latencies from different items are assumed to be independent. This yields the test information

$$I(\theta, \omega) = \sum_{i=1}^{n} \begin{bmatrix} I_{xi}(\theta) & 0 \\ 0 & 0 \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} I_{ti[11]}(\theta, \omega) & I_{ti[12]}(\theta, \omega) \\ I_{ti[12]}(\theta, \omega) & I_{ti[22]}(\theta, \omega) \end{bmatrix}. \tag{6}$$

Element $I_{t[12]}(\theta, \omega)$ and element $I_{t[22]}(\theta, \omega)$ represent the other entries in the information matrix of a single response latency. These elements could be denoted as the information of a single response latency about general response speed ($I_{t[22]}(\theta, \omega)$) and the joint information about the target trait and general response speed ($I_{t[12]}(\theta, \omega)$). As these elements are not needed in the following discussion, they will not be given in formulas. The second summand in Equation 6 represents the gain in information that results from the consideration of the response latencies. As can be seen, the consideration of the response latencies increases the information of the test. Additionally, by arguments similar to those of Sorensen and Gianola (2002, p. 182), it can be shown that the variance of the trait estimator $\hat{\theta}$ is reduced.

Although the preceding arguments were based on the assumption of log-normally distributed response latencies, they are more general than they might appear. In fact, they are not bounded to specific distributional assumptions as long as Equation 1 holds. This can be shown by applying findings for quasi–maximum likelihood estimation (Wedderburn, 1974).

Method

Participants

The model was estimated using data collected by Ortner (2008) consisting of 171 men called up for military service. If they agreed (about 80%), they were given a personality questionnaire after the standardized psychological testing conducted by the Psychological Service of the Austrian Armed Forces. Persons were included only if they were evaluated as being motivated by the conductor, if they had solved the standardized battery quickly, and if no language problems were known. To reduce faking, the conductor pointed out that all results would be handled anonymously and would not be evaluated for the military appropriateness of the persons. One individual was removed from data analysis because of unusually fast response latencies indicating low motivation.

Measures

We used seven scales from the German pretest version of the Eysenck Personality Profiler, which had previously been shown to fit the dichotomous Rasch model. The entire questionnaire was published with a reduced number of items by Eysenck, Wilson, and Jackson (1998). It measures 21 personality traits that are consistent with the three major dimensions of personality as defined by Eysenck.

Items were presented on 20 computers that were connected to a server but were not connected to the Internet. To minimize the start-up time of each computer, pictures were saved in each computer's local storage. As a quality control of the measurement,

from time to time the Psychological Service of the Austrian Armed Forces manually recorded the system's response latencies and compared the measures with the automatic recordings. The recordings showed that on average only 1/100 seconds of variation had to be taken into account at the item level.

Procedure

Scales were analyzed separately. Data analysis consisted of four steps. In the first step, a Rasch model was fit to the responses. Item parameters were estimated according to the marginal maximum likelihood principle with the Rasch procedure of the R program (Rizopoulos, 2006). Model fit was verified by using Pearson's χ^2 statistic to compare the number of expected and observed response patterns. As the expected frequencies were very low, the distribution of the χ^2 statistic could not be approximated by the χ^2 distribution. Therefore, the p value was approximated by means of a parametric bootstrapping method. Simulations have demonstrated the appropriateness of this approach (von Davier, 1997). Additionally, the deviations of observed and expected cell counts were inspected for the $n(n - 1)$ two-way margin tables. As a rule of thumb, residuals greater than 4.0 can be considered indicators of model misfit (Bartholomew & Knott, 1999). The fitted Rasch model allowed the estimation of the probabilities of the given responses by using the estimated item parameters and the empirical Bayes estimates of the trait levels.

The second step consisted of linear regression analysis. The log response latency for each item was predicted from the estimated response probability. In this step, we investigated whether there was a linear relation between the response probability and the expected log response latency. Fit of the linear regression models was assessed with residual plots: Variance homogeneity and linearity were evaluated by plotting standardized residuals against fitted values. Normality was checked by a Q–Q plot. Only items with acceptable model fit and a significant relation between their response probability and log response latency were analyzed further. In the third step, a multilevel model was fit to the log response latencies of the selected items. According to Equation (1), the following predictors were included: dummy variables for the time demand of an item, the response probability, and interaction effects to allow for item-specific effects of the response probability. Additionally, a random intercept was included for every individual to account for individual differences in respond speed. Contrary to the single-regression analysis, the multilevel model permitted the decomposition of the residual variance into two parts: one part due to individual differences in response speed and an unsystematic part due to pure error. In the fourth step, the information of the test was estimated using the results from the multilevel results model.

Having fit the Rasch model, the bootstrap fit tests revealed that only three scales out of seven were consistent with the Rasch model, namely, the anxious–calm scale (MBES), the manipulative–empathic scale (MEGO), and the depressed–happy scale (MSPO). The p values of the bootstrap tests were $p = .48$ for the MBES scale, $p = .70$ for the MEGO scale, and $p = .30$ for the MSPO scale. Additionally, the

Table 1. Item Parameters of the Items: Difficulty α_0, Discrimination α_1, and Standard Error of Estimation

Scale	Item	α_0	SE	α_1	SE	Item	α_0	SE	α_1	SE
MBES	1	−1.79	0.24	1.20	0.11	9	−2.43	0.28	1.20	0.11
	2	−1.79	0.24	1.20	0.11	10	−2.36	0.28	1.20	0.11
	3	−1.37	0.22	1.20	0.11	11	−3.24	0.35	1.20	0.11
	4	−3.04	0.32	1.20	0.11	12	−1.45	0.22	1.20	0.11
	5	−1.93	0.25	1.20	0.11	13	−1.62	0.23	1.20	0.11
	6	−1.30	0.22	1.20	0.11	14	−0.87	0.21	1.20	0.11
	7	−1.75	0.24	1.20	0.11	15	−2.50	0.28	1.20	0.11
	8	−1.79	0.24	1.20	0.11					
MEGO	1	−1.40	0.20	0.68	0.08	9	−0.44	0.17	0.68	0.08
	2	−0.47	0.17	0.68	0.08	10	−0.15	0.17	0.68	0.08
	3	−1.51	0.21	0.68	0.08	11	−1.55	0.21	0.68	0.08
	4	0.50	0.18	0.68	0.08	12	−2.97	0.37	0.68	0.08
	5	−0.66	0.20	0.68	0.08	13	0.33	0.17	0.68	0.08
	6	−1.29	0.20	0.68	0.08	14	0.05	0.17	0.68	0.08
	7	−1.59	0.21	0.68	0.08					
	8	−0.87	0.18	0.68	0.08					
MSPO	1	−1.22	0.20	0.93	0.08	9	−1.12	0.20	0.93	0.08
	2	0.20	0.18	0.93	0.08	10	−0.29	0.18	0.93	0.08
	3	−0.95	0.19	0.93	0.08	11	−2.26	0.26	0.93	0.08
	4	1.38	0.21	0.93	0.08	12	−1.02	0.20	0.93	0.08
	5	−0.82	0.19	0.93	0.08	13	1.10	0.20	0.93	0.08
	6	0.63	0.19	0.93	0.08	14	−0.58	0.19	0.93	0.08
	7	−1.33	0.21	0.93	0.08	15	−0.35	0.18	0.93	0.08
	8	−1.26	0.20	0.93	0.08					

Note: MBES = anxious–calm scale; MEGO = the manipulative–empathic scale; MSPO = depressed–happy scale.

comparison of expected and observed cell counts in the two-way margin tables did not show any evidence of misfit. Therefore, the description of the results was limited to these three scales. The item parameters of the items can be found in Table 1. These item parameters imply reliabilities of $r_{xx} = .75$ for the MBES scale, $r_{xx} = .54$ for the MEGO scale, and $r_{xx} = .70$ for the MSPO scale (Kiefer & Wolfowitz, 2010).

   After estimation of the item parameters, the individual trait levels were estimated. This allowed the calculation of the probability of the given response for every item and every individual according to Equation 2. The estimated probability was used as a predictor of the log response latencies in a linear regression analysis. During this stage of analysis, no effort was made to decompose the unexplained part of the log response latency into a pure error term and a part due to the response speed of an individual. The results of the different regression analyses can be found in Table 2. As predicted, the response probability was negatively related to the log response latency for almost every item. However, not every relation was significant. In fact, only 17 of the 44 regression coefficients could be considered different from zero

Table 2. Regressions Analysis of Response Probability on Log Response Latency: Coefficient, Standard Error, and Significance Level (Two Sided)

Scale	Item	β_1	SE	$P(>\|t\|)$	Item	β_1	SE	$P(>\|t\|)$
MBES	1	−0.095	0.159	0.55	9	−0.258	0.127	0.04
	2	−0.060	0.129	0.64	10	−0.286	0.157	0.07
	3	−0.187	0.164	0.26	11	−0.216	0.148	0.15
	4	−0.226	0.176	0.20	12	−0.397	0.129	0.00
	5	−0.381	0.146	0.01	13	−0.128	0.160	0.43
	6	−0.418	0.147	0.01	14	−0.289	0.146	0.05
	7	−0.261	0.140	0.07	15	−0.319	0.134	0.02
	8	−0.565	0.164	0.00				
MEGO	1	0.053	0.152	0.73	9	−0.003	0.223	0.99
	2	−0.594	0.254	0.02	10	0.056	0.355	0.88
	3	−0.446	0.171	0.01	11	0.017	0.140	0.90
	4	−0.500	0.244	0.04	12	0.078	0.147	0.59
	5	−0.213	0.206	0.30	13	−0.692	0.251	0.01
	6	−0.254	0.149	0.09	14	−0.061	0.343	0.86
	7	−0.064	0.127	0.62				
	8	−0.004	0.157	0.98				
MSPO	1	−0.279	0.130	0.03	9	−0.339	0.120	0.01
	2	−0.474	0.199	0.02	10	−0.217	0.184	0.24
	3	−0.151	0.165	0.36	11	−0.244	0.138	0.08
	4	−0.618	0.129	0.00	12	−0.158	0.172	0.36
	5	−0.311	0.192	0.11	13	−0.251	0.165	0.13
	6	−0.678	0.203	0.00	14	0.006	0.199	0.98
	7	−0.058	0.169	0.73	15	−0.271	0.186	0.15
	8	−0.539	0.160	0.00				

Note: MBES = anxious–calm scale; MEGO = the manipulative–empathic scale; MSPO = depressed–happy scale.

(or 21 for one-sided testing). This might be due to the fact that the effect of the response probability overlapped with individual differences in response speed. Therefore, it was not surprising that the explained variance of response time was rather low, ranging between $R^2 = .00$ and $R^2 = .12$.

To check the assumption of the linearity of the regression analyses, standardized residuals were plotted against fitted values. None of the plots revealed evidence of violations of this assumption. Additionally, the assumption of log-normally distributed response latencies was investigated. This consisted of checking for whether the regression residuals were normally distributed. Some items showed an excellent fit to the log-normal distribution. For example, the Q–Q plot of Item 6 of the MBES scale is given in Figure 1.

In the following, only the MBES scale and the MSPO scale were analyzed further. The MEGO scale was excluded due to the low discrimination parameters of the items, calling into question its construct validity. New scales were constructed by selecting
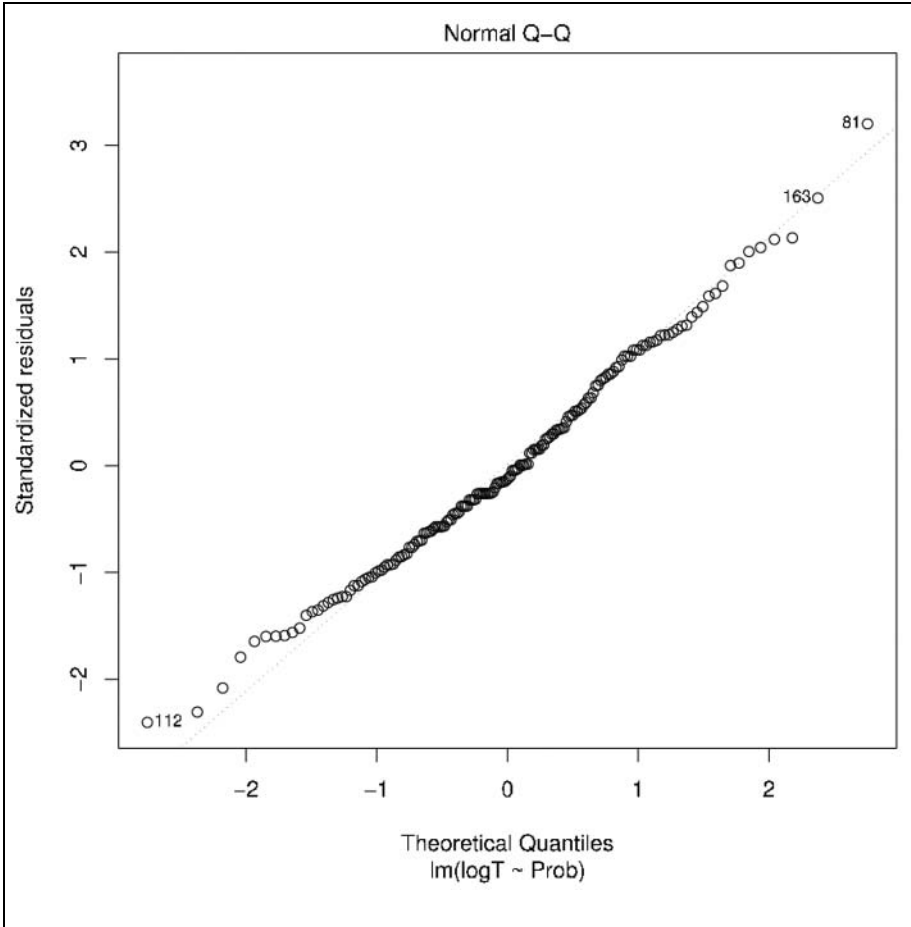
Figure 1. *Q–Q* plot: Fit of log-normal distribution to Item 6 of the MBES scale

items with good model fit and a strong relation between response probability and log response latency. This resulted in choosing Items 5, 6, 7, 8, 10, and 12 of the MBES scale and Items 2, 6, 8, 9, 10, and 11 of the MSPO scale. The new scales were analyzed with a multilevel model. To account for the dependency of response latencies within the same subject, a random intercept was included. This random intercept was supposed to reflect the general response speed ω of an individual. In the first analysis, three sets of predictors with fixed effects were included, namely, dummy variables for the different items, the response probability, and interaction effects between the item and the response probability. As none of the interaction effects was significant at $\alpha = .05$, they were removed from the model. For the MBES scale, the effect of the response probability was estimated as $\beta_1 = -0.29$ ($p < .01$). The standard deviation of the residuals was estimated as $\sigma_e = 0.38$, whereas the standard deviation of

the random intercept was estimated as $\sigma_\omega = 0.24$. For the MSPO scale, the resulting estimates were $\beta_1 = -0.32$ ($p < .01$) for the response probability, $\sigma_e = 0.36$ for the standard deviation of the residuals, and $\sigma_\omega = 0.21$ for the standard deviation of the intercept. This confirms that log response latency is related to the response probability.

Based on these estimates, it was possible to calculate the information of the test when considering responses and response latencies (see Equations 4 and 5). This yielded the test information curves given in Figure 2.

As can be seen in Figure 2, both information curves were peaked. Because of the consideration of response latencies, the test information about θ increases maximally by 13% for the MBES scale and by 17% for the MSPO scale. This shows that a test gains in precision when considering response latencies.

Discussion

Based on the preceding literature review indicating the informative value of item response latencies for personality assessment, the present study employed a new IRT approach that can be used to model and use response latencies in personality questionnaires. The previously discovered inverted-U effect linking trait and response latency was taken as a starting point for modeling the response latencies. Different from the previous studies that only ascertained this relationship, a detailed model was set up for response latencies linking them to the probability of the given response. This provided the possibility of assigning individuals to certain trait levels given their response latencies, allowing the incorporation of the response latencies into the diagnostic assessment of individuals.

Beside the theoretical and descriptive value of the presented model, our data confirmed that the quantities—response latency and response probability—were empirically related. However, with reference to the analyzed items in our study, not all of them provided evidence for a strong relationship between response latency and response probability. There are several reasons that can account for this lack of evidence: First, it has been shown that certain item characteristics increase response latencies independent of trait characteristics. It even seems possible that certain item features suppress or even replace the effects of individual trait characteristics, such as item length and emotional evocative character (Dunn et al., 1972; Neubauer & Malle, 1997; Temple & Geisinger, 1990; Tyron & Mulloy, 1993). Second, the tests that we used were not constructed and optimized for the use of response latencies as trait indicators because this criterion had not been considered during test development. And third, the relation between the response probability and the response latency might be underestimated. The estimation of individual response probabilities depends crucially on the estimated trait levels and the estimates of the item parameters. As both quantities are only estimated and do not coincide with the true values exactly, the employed probability values are contaminated with measurement error. It is well known that the strength of linear relations is underestimated when predictors contain measurement error. Additionally, future studies may also include indicators of further
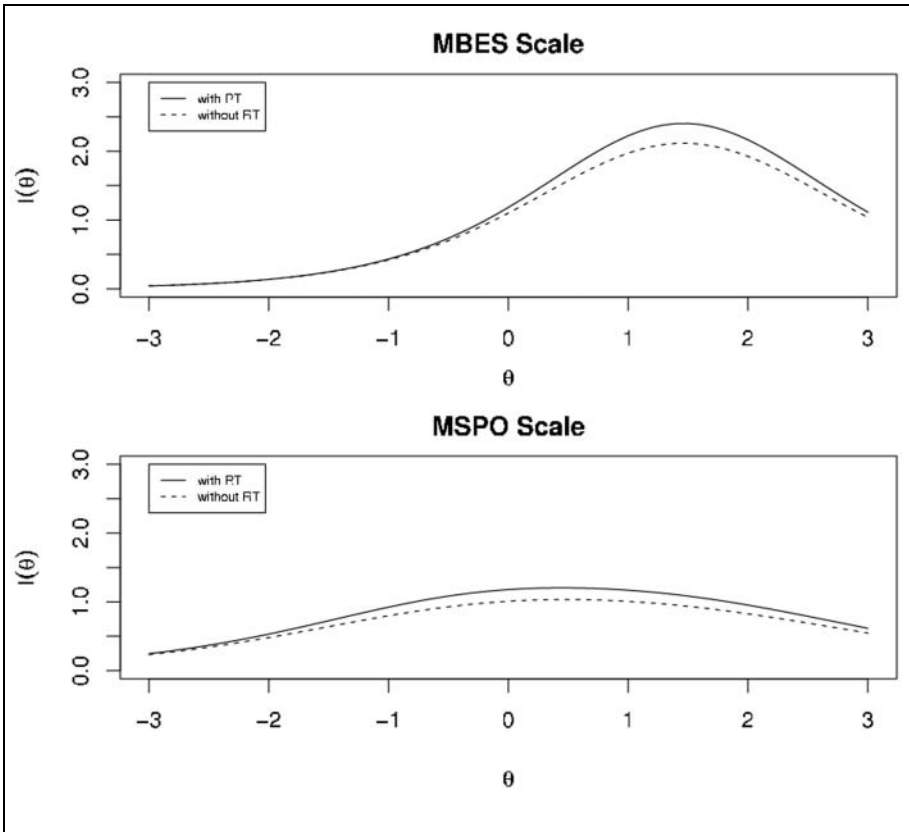
Figure 2. Test information when ignoring and including response times

test-taker characteristics such as the individual disposition to answer honestly or to fake (Holden, 1995; Holden & Hibbs, 1995), a person's reading abilities (Jansen, 1997), impulsivity (Vigil & Codorniu, 2004), psychomotor reaction speed (Genser, 1988), and motivation (Callegaro, Yang, Bhola, & Dillman, 2005). These characteristics can potentially be used to explain a larger amount of variance.

   In this study, the practical usability as well as the utility of the presented model was demonstrated. The additional inclusion of response latencies increased the test information maximally by 17%. This increase can be interpreted in two ways: Seeing response latencies in computerized item presentation as information almost free of cost, the results—as any additional information—provide an economic way to increase individuals' trait information through a testing procedure constant in time. When evaluating the presented findings, it has to be taken into account that the realized increase of information does not reflect the true potential of response latencies for trait estimation. As mentioned above, as a first reason, items were not designed and proven for response time–based assessment. And second, the increase of information

was underestimated because of the underestimation of the relation between response probability and response latency.

In line with these considerations, future studies should address confounding effects of certain item characteristics and trait-based latencies. Furthermore, it should be a future aim of research to find out whether the theory of schema-driven response latencies can be similarly applied and generalized across different traits. Although previous studies have already included a range of different traits (Akrami et al., 2007; Amelang et al., 1991; Holden & Fekken, 1993), future studies might identify more favorable traits, item formats, and/or item contents to maximize trait information given in response latencies. And finally, more appropriate strategies are required for calculation of response probabilities for this purpose in the future.

Overall, the present study enriches the existing methodological view and provides additional tools for employing response latencies for the purpose of psychological assessment. For applied situations, our study design may serve as a first example for how to practically incorporate response latencies into a given testing procedure. Our data suggest that in the future, response latencies as a source of information should be taken into account for economic estimation of an individual's trait level.

Declaration of Conflicting Interests

Funding

References

Akrami, N., Hedlund, L., & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: The inverted-U effect revisited. *Personality and Individual Differences, 43*, 611-618.

Amelang, M., Eisenhut, K., & Rindermann, H. (1991). Responding to adjective check list items: A reaction time analysis. *Personality and Individual Differences, 12*, 523-533.

Anastasi, A. (1976). *Psychological testing*. New York, NY: Macmillan.

Austin, E. (2009). A reaction time study of responses to trait and ability emotional intelligence test items. *Personality and Individual Differences, 46*, 381-383.

Baker, F., & Seock-Ho, K. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.

Bartholomew, D., & Knott, M. (1999). *Latent variable models and factor analysis*. London, England: Arnold.

Birnbaum, A. (1968). Classification by ability levels. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores* (pp. 436-452). Reading, MA: Addison-Wesley.

Callegaro, M., Yang, Y., Bhola, D., & Dillman, D. (2005). *Response latency as an indicator of optimization: A study comparing job applicants' and job incumbents' response time on a web survey* (Survey Research and Methodology No. 11). Lincoln: University of Nebraska.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*, 63-65.

Costa, P., & McCrae, R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Dunn, T., Lushene, R., & O'Neil, H. (1972). Complete automation of the MMPI and a study of its response latencies. *Journal of Consulting and Clinical Psychology, 39*, 381-387.

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

Eysenck, H., Wilson, C., & Jackson, C. (1998). *Eysenck Personality Profiler* (EPP-D). Frankfurt, Germany: Swets.

Fekken, G., & Holden, R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality, 26*, 103-120.

Ferrando, P., & Lorenzo-Sevas, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement, 31*, 525-543.

Genser, S. (1988). Computer response time measurement of mood, fatigue and symptom scale items: Implications for scale response time uses. *Computers in Human Behavior, 4*, 95-109.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Höhn, E. (1959). Theoretische Aspekte der Inhaltsanalyse projektiver Tests. *Psychologische Forschung, 26*, 13-74.

Holden, R. (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science, 27*, 343-355.

Holden, R., & Fekken, G. (1990). Structured psychopathological test item characteristics and validity. *Psychological Assessment, 2*, 35-40.

Holden, R., & Fekken, G. (1993). Can personality test item response latencies have construct validity? Issues of reliability and convergent and discriminant validity. *Personality and Individual Differences, 15*, 243-248.

Holden, R., Fekken, G., & Cotton, D. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment, 3*, 111-118.

Holden, R., Fekken, G., & Jackson, D. (1985). Structured personality test item characteristics and validity. *Journal of Research in Personality, 19*, 386-394.

Holden, R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality, 29*, 362-372.

Holden, R., & Kroner, D. (1992). The relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment, 4*, 170-173.

Jackson, D. (1989). *The Basic Personality Inventory manual*. Port Huron, MI: Research Psychologists Press.

Jackson, D. (1994). *Jackson Personality Inventory: Revised manual*. Port Huron, MI: Research Psychologists Press.

Jansen, M. (1997). Rasch's model for reading speed with manifest explanatory variables. *Psychometrika, 62*, 393-409.

Kiefer, J., & Wolfowitz, J. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Educational Review, 11*, 179-188.

Klein, J., & Moeschberger, M. (1997). *Survival analysis: Techniques for censored and trun-cated data*. Berlin, Germany: Springer.

Kuiper, N. (1981). Convergent evidence for the self as a prototype: The ''Inverted-U RT Effect'' for self and other judgements. *Personality and Social Psychology Bulletin, 7*, 438-443.

Lavergne, C., & Vigneau, F. (1997). Response speed on aptitude tests as an index of intellectual performance: A developmental perspective. *Personality and Individual Differences, 23*, 283-290.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

MacLennan, R., Jackson, D., & Bellantino, N. (1988). Response latencies and the computerized assessment of intelligence. *Personality and Individual Differences, 9*, 811-816.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*, 63-78.

Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy, 23*, 563-569.

Neubauer, A., & Malle, B. (1997). Questionnaire response latencies: Implications for personality assessment and self-schema theory. *European Journal of Psychological Assessment, 13*, 109-117.

Ortner, T. M. (2008). Effects of changed item order: A cautionary note to practitioners on jumping to computerized adaptive testing for personality assessment. *International Journal of Selection and Assessment, 16*, 249-257.

Popham, S., & Holden, R. (1990). Assessing MMPI constructs through the measurement of response latencies. *Journal of Personality Assessment, 54*, 469-478.

Rafaeli, S., & Tractinsky, N. (1991). Time in computerized tests: A multitrait, multimethod investigation of general-knowledge and mathematical-reasoning on-line examinations. *Computers in Human Behavior, 7*, 215-225.

Rizopoulos, D. (2006). Itm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1-25.

Schnipke, D., & Scrams, D. (2002). Exploring issues of examinee behavior: Insights gained from response-time analysis. In C. Mills, M. Potenza, J. Fremer, & W. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 237-266). Mahwah, NJ: Lawrence Erlbaum.

Sorensen, D., & Gianola, D. (2002). *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. Berlin, Germany: Springer.

Stricker, L., & Alderton, D. (1999). Using response latency measures for a biographical inventory. *Military Psychology, 11*, 169-188.

Stroop, R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Temple, D. E., & Geisinger, K. F. (1990). Response latency to computer-administered inventory items as an indicator of emotional arousal. *Journal of Personality Assessment, 54*, 289-297.

Thissen, D. (1983). Timed testing: An approach using item response theory. In D. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 179-203). New York, NY: Academic Press.

Tyron, W., & Mulloy, J. (1993). Further validation of computer-assessed response time to emotionally evocative stimuli. *Journal of Personality Assessment, 61*, 231-236.

van der Linden, W. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*, 181-204.

van der Linden, W. (2008). Using response times for item selection in adaptive tests. *Journal of Educational and Behavioral Statistics, 31*, 5-20.

Vigil, A., & Codorniu, M. (2004). Aggression and inhibition deficits, the role of functional and dysfunctional impulsivity. *Personality and Individual Differences, 37*, 1431-1440.

von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research Online*, 2.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models and the Gauss–Newton method. *Biometrika, 61*, 439-447.