T E C H N I C A L
R E P O R T

0689

# A SPEEDED ITEM RESPONSE MODEL WITH GRADUAL PROCESS CHANGE

GOEGEBEUR, Y., DE BOECK, P., WOLLACK, J.A. and A.S. COHEN

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

# A speeded item response model with gradual process change

Yuri Goegebeur *
Paul De Boeck ‡
James A. Wollack §
Allan S. Cohen ¶

**Abstract**

An item response theory model for dealing with test speededness is proposed. The model consists of two random processes, a problem solving process and a random guessing process, with the random guessing gradually taking over from the problem solving process. The involved change point and change rate are considered random parameters in order to model examinee differences in both respects. The proposed model is evaluated on simulated data and in a case study.
**Key words:** Rasch model, local item dependence, test speededness.

*Department of Psychology, K.U.Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: yuri.goegebeur@psy.kuleuven.be

‡Department of Psychology, K.U.Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: paul.deboeck@psy.kuleuven.be

§School of Education, Testing and Evaluation Services, University of Wisconsin - Madison, 373 Educational Sciences, 1025 W Johnson St, Madison, WI 53706. Email: jwollack@wisc.edu

¶Department of Educational Psychology, University of Georgia, 329 Aderhold Hall, Athens, GA 30606. Email: acohen@uga.edu

# 1 Introduction

Test speededness has been modeled using two alternative item response theory (IRT) approaches both of which assume a single point at which the examinee's response strategy switches to an alternative response strategy due to time limits being reached for the test (Bolt *et al.*, 2001, Yamamoto, 1987 and Yamamoto and Everson, 1997). In this paper, we propose an alternative model in which the response strategies switch more gradually. We show that this alternative model is in fact a general case which subsumes both previous explanations of speededness and provides a more realistic view of test speededness. In addition, this alternative model provides an opportunity to consider modeling other psychological processes, particularly ones which may change gradually, such as learning or change in attitudes or preferences.

Let $Y_{pi}$ denote the binary response (incorrect/correct, coded $Y_{pi} = 0$ and $Y_{pi} = 1$, respectively) of examinee $p$, $p = 1, \ldots, P$, to item $i$, $i = 1, \ldots, I$. In the classical one-parameter Rasch model (1PL) (Rasch, 1960) $Y_{pi}$ depends on the examinee ability $\theta_p$ and item difficulty $\beta_i$ in the following way

$$Y_{pi}|\theta_p \quad \sim \quad \text{Bern}(P_i(\theta_p))$$

with

$$P_i(\theta_p) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \tag{1}$$

and $\theta_p \sim N(0, \sigma_\theta^2)$ if the marginal maximum likelihood formulation is chosen. Moreover, conditional on $\theta_p$ all responses of subject $p$ are assumed independent; this is the so-called local item independence condition. Formally, denoting $\mathbf{Y}'_p = (Y_{p1}, \ldots, Y_{pI})$,

$$P(\mathbf{Y}_p = \mathbf{y}_p | \ \theta_p) = \prod_{i=1}^{I} [P_i(\theta_p)]^{y_{pi}} [1 - P_i(\theta_p)]^{1-y_{pi}}.$$

The Rasch model has been extended in several ways. In the two-parameter logistic model (2PL) (Birnbaum, 1968) the difference $\theta_p - \beta_i$ is weighted by an item discrimination parameter $\alpha_i$:

$$P_i(\theta_p) = \frac{\exp(\alpha_i(\theta_p - \beta_i))}{1 + \exp(\alpha_i(\theta_p - \beta_i))}, \tag{2}$$

so that the influence of examinee ability on outcome depends on the item. The three-parameter logistic model (3PL) (Birnbaum, 1968) extends the 2PL with an item specific guessing parameter $c_i$:

$$P_i(\theta_p) = c_i + (1 - c_i)\frac{\exp(\alpha_i(\theta_p - \beta_i))}{1 + \exp(\alpha_i(\theta_p - \beta_i))}.$$

The parameter $c_i$ clearly reflects the probability of a correct answer under random guessing. For further interpretations of the 3PL, we refer to Hutchinson (1991).

The Rasch model, and item response theory models in general, are not robust with respect to violations of the local item independence assumption. The inclusion of items with local item dependence may result in contaminated estimates of test reliability, person and item parameters, standard errors and equating coefficients, see for instance Yen (1984), Thissen *et al.* (1989), Sireci *et al.* (1991), Yen (1993), Wainer and Thissen (1996), Lee *et al.* (2001) and Tuerlinckx and De Boeck (2001).

Yen (1993) and Ferrara *et al.* (1999) provide a detailed taxonomy of possible reasons for the existence of local item dependency. One of the most prevalent causes in educational testing is test speededness. Test speededness refers to testing situations in which some examinees do not have ample time to answer all questions. Speededness effects are often detrimental to the intended functioning of the test in that the speed with which one responds is usually not an important part of the construct of interest. Examinees affected by test speededness hurry through, randomly guess on or even fail to complete items, usually at the end of the test, and hence receive ability estimates that may underestimate their capacities. On the other hand, the item difficulty parameters of items administered late in the test tend to be overestimated (Douglas *et al.*, 1998 and Oshima, 1994).

Item response theory models dealing with test speededness are relatively new.

The hybrid model of Yamamoto and Everson (1997) uses multiple item response theory models to describe the behavior of examinees. A classical item response model is valid throughout most of the test but end-of-test items are answered randomly by some subset of examinees. The model identifies $M$ possible latent classes, one for whom an item response model is valid for all items, and $M-1$ classes with an item response model describing answers to the first $I-m$ items and random guessing on the last $m$ items, $m = 1, \ldots, M-1$. Formally,

$$P_i^{(m)}(\theta_p^{(m)}) = \begin{cases} \frac{\exp(\alpha_i(\theta_p^{(m)} - \beta_i))}{1+\exp(\alpha_i(\theta_p^{(m)} - \beta_i))}, & i \leq I - m, \\ c_i, & i > I - m, \end{cases}$$

with $m = 0, \ldots, M-1$. Clearly, speededness is unlikely to be so straightforward, as students do not switch immediately to random guessing beyond some point.

Bolt *et al.* (2002) extend the mixture Rasch model proposed by Rost (1990) to distinguish latent classes of examinees according to the existence of speededness in their item response patterns. Ordinal constraints are imposed on the item difficulty parameters across classes so as to distinguish a class having no speededness effects from a class whose responses are affected by speededness. In particular, for items early in the test, the item difficulty parameters are constrained to be equal in the two classes; however, the item difficulty parameters of end-of-test items in the speeded class are constrained to be larger than the respective item difficulty parameters in the nonspeeded class. Let $g$ denote a class indicator with $g = 0, 1$ referring to the nonspeeded and speeded class, respectively, and let $k$ denote the first item where the examinees experience the effects of test speededness. The mixture Rasch model can then be stated as

$$P_i^{(g)}(\theta_p^{(g)}) = \frac{\exp(\theta_p^{(g)} - \beta_i^{(g)})}{1 + \exp(\theta_p^{(g)} - \beta_i^{(g)})},$$

with

$$\begin{aligned} \beta_i^{(0)} &= \beta_i^{(1)} \quad \text{for} \quad i < k, \\ \beta_i^{(0)} &< \beta_i^{(1)} \quad \text{for} \quad i \geq k. \end{aligned}$$

The item difficulty estimates obtained in the nonspeeded class provide more suitable estimates of the Rasch difficulties of end-of-test items than the difficulties estimated using all examinees. Although this model has worked quite well at identifying test speededness, it does not allow for different examinees becoming speeded at different points in the test. Since such differences are plausible, in this paper, we propose a model that provides for this kind of transition as a random effect within examinees.

The remainder of this paper is organized as follows. In the next section we propose an item response model that accommodates the disadvantages of the hybrid model and the mixture Rasch model. The model can be seen as consisting of two random processes, a problem solving process and a random guessing process, with the random guessing gradually taking over from the problem solving process. In this paper we use a Rasch process for the problem solving component of the model. The involved change point and change rate are considered random parameters in order to model examinee differences in both respects. The model was first formulated by Wollack and Cohen (2004) as a model to simulate speededness data, but it will be treated here as a full-fledged model for test data which can also be estimated. In Section 3 we evaluate the performance of the model on the basis of a simulation study. The final section reports the results of applying the model to a mathematics placement test.

## 2 A model for speeded test data with gradual process change

In this section we propose a new item response model for dealing with speeded test data. Under the model, responses to items early in the test are governed by a Rasch model. Beyond some point the success probability gradually decreases and eventually reduces to the success probability under random guessing. Both change point and change rate are examinee specific.

Using the same notation as before, the model can be stated as

$$Y_{pi} | \theta_p, \eta_p, \lambda_p \sim \text{Bern}(\pi_{pi})$$

with

$$\pi_{pi} = c_i + (1 - c_i) P_i(\theta_p) \min \left\{ 1, \left[ 1 - \left( \frac{i}{I} - \eta_p \right) \right]^{\lambda_p} \right\}, \qquad (3)$$

where $P_i(\theta_p)$ is given by (1) or (2), $\eta_p$ ($\eta_p \in [0,1]$) represents the speededness point and $\lambda_p$ ($\lambda_p \geq 0$) the speededness rate of examinee $p$. The speededness point parameter $\eta_p$ identifies the point in the test, expressed as a fraction of the number of items, where examinee $p$ first experiences an effect due to speeding. For items with $i \leq \eta_p I$ there is no effect of speeding. Once the examinee passes his/her speededness point, $i/I - \eta_p$ is positive, resulting in a decrease of $\pi_{pi}$. The rate of decrease of $\pi_{pi}$ is controlled by the parameter $\lambda_p$, with larger $\lambda_p$ values resulting in a faster decrease. In Figure 1 we illustrate the role of $\eta$ and $\lambda$ by plotting the decay function $\min\{1, [1-(x-\eta)]^{\lambda}\}$ for some values of $\eta$ and $\lambda$.

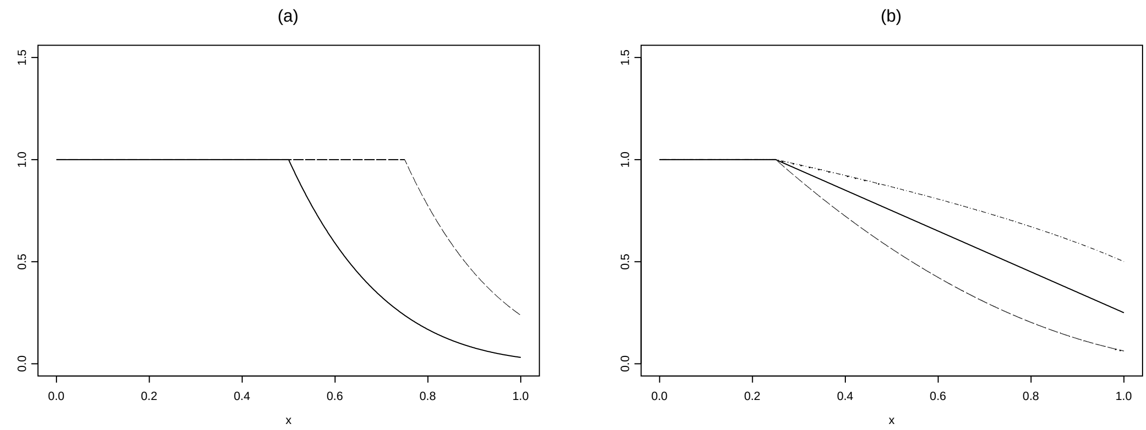


Figure 1: (a) $\min\{1, [1-(x-\eta)]^{\lambda}\}$ for $\lambda = 5$, $\eta = 0.5$ (solid line) and $\eta = 0.75$ (broken line), (b) $\min\{1, [1-(x-\eta)]^{\lambda}\}$ for $\eta = 0.25$, $\lambda = 1$ (solid line), $\lambda = 2$ (broken line) and $\lambda = 0.5$ (broken-dotted line).

The rationale for the proposed model is as follows. Denote $P_i(\eta_p, \lambda_p) = \min\{1, [1-(i/I-\eta_p)]^{\lambda_p}\}$. When examinee $p$ encounters item $i$, he/she answers according to either a Rasch process or a random guessing process, with probabilities $P_i(\eta_p, \lambda_p)$ and $1 - P_i(\eta_p, \lambda_p)$ respectively. Under random guessing the answer is correct with probability $c_i$. Under the Rasch process the examinee knows the answer with probability $P_i(\theta_p)$; if ignorant the examinee guesses at random. In Figure 2 we visualize the model with a decision tree. Clearly,

$$P(Y_{pi} = 1|\theta_p, \eta_p, \lambda_p) = P_i(\eta_p, \lambda_p)P_i(\theta_p) + P_i(\eta_p, \lambda_p)[1 - P_i(\theta_p)]c_i + [1 - P_i(\eta_p, \lambda_p)]c_i,$$

which simplifies to (3).

Model (3) has some interesting limiting cases:

- if $[1-(i/I-\eta)]^{\lambda} = 0$ for $i/I > \eta$ (this corresponds to the limiting case $\lambda \to +\infty$), then (3) reduces to one of the speeded classes in the hybrid model, and speededness is modeled as random guessing,

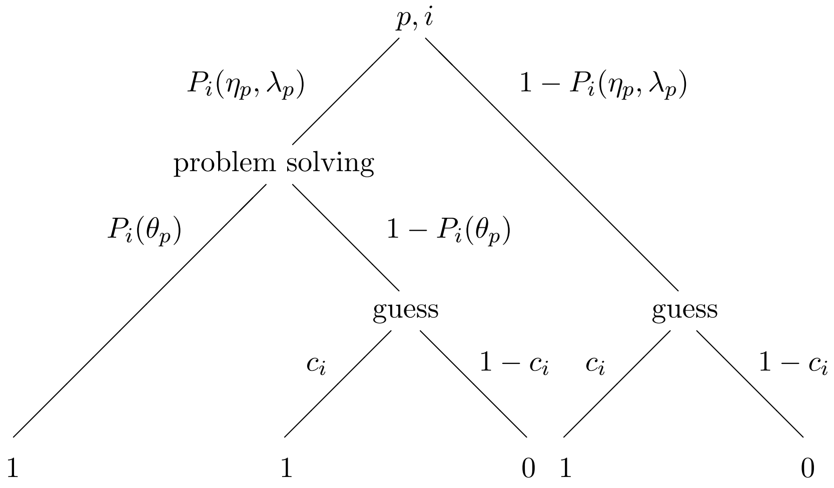


Figure 2: Decision tree representation of speededness model.

- in case $\lambda = 0$ or $\eta = 1$, the proposed model reduces to 1PL extended with random guessing or 3PL,
- in case $\eta = 0$ and $\lambda > 0$ the examinee guesses at random at least to some degree from the first item up to the final item,
- as with the 3PL, $c_i$ is the lower asymptote for $\theta \to -\infty$.

Note that the 3PL model is obtained as a limiting case for $\lambda = 0$ (whatever the value of $\eta$) or for $\eta = 1$ (whatever the value of $\lambda$) and hence is not uniquely identified within the proposed test speededness model. This of course may entail estimation difficulties, such as non-convergence of the optimization algorithm or ill-conditioned observed information matrices, when model (3) is fitted to data that are not affected by test speededness. We will come back to this issue in Section 3.

As is usual in item response theory, the person ability parameter is assumed to be normally distributed with mean zero and variance $\sigma_\theta^2$. Concerning the parameters $\eta_p$ and $\lambda_p$ we make, without loss of generality, the following distributional assumptions:

$$\begin{aligned} \eta_p &\sim \text{Beta}(\alpha, \beta), \\ \lambda_p &\sim \log N(\mu_\lambda, \sigma_\lambda^2). \end{aligned}$$

For estimation, we restrict the discussion to the marginal maximum likelihood method. If the model of interest is given by (3)-(1) with a common unknown random guessing parameter $c$, then the parameters to be estimated are $(\beta_1, \ldots, \beta_I, c, \sigma_\theta^2, \alpha, \beta, \mu_\lambda, \sigma_\lambda^2)$, whereas under (3)-(2) the parameters to be estimated are $(\alpha_1, \ldots, \alpha_I, \beta_1, \ldots, \beta_I, c, \alpha, \beta, \mu_\lambda, \sigma_\lambda^2)$. In the latter case $\sigma_\theta^2$

has to be fixed at some positive constant for identification purposes. For convenience the vector of unknown parameters will be denoted by $\boldsymbol{\xi}$. In the marginal maximum likelihood method the random effects are integrated out and the resulting likelihood is maximized with respect to the unknown parameters. Under (3) and denoting joint density function of $\theta_p$, $\eta_p$ and $\lambda_p$ by $g$ the marginal likelihood function is simply

$$L(\boldsymbol{\xi}) = \prod_{p=1}^{P} \int_{\mathbb{R}} \int_{0}^{1} \int_{0}^{\infty} \prod_{i=1}^{I} P(Y_{pi} = y_{pi}|\theta_p, \eta_p, \lambda_p) g(\theta_p, \eta_p, \lambda_p) d\lambda_p d\eta_p d\theta_p. \tag{4}$$

The integrals involved in (4) can be numerically approximated by a quadrature method and the optimization can be performed using a standard Newton-Raphson algorithm. The SAS NLMIXED procedure fits nonlinear mixed models with multivariate normal random effect distributions. However, as long as $g$ in (4) is characterized by a normal dependence structure (copula) NLMIXED can be used to fit model (3), whatever the functional form of the (continuous) marginal random effect distribution functions. Indeed, as shown in Proposition 1 (see Appendix 2), in case of a normal dependence function, appropriately chosen compositions of probability integral transforms and inverse probability integral transforms of the marginal distributions yield a multivariate normal distribution for the transformed random effects. As an alternative to the SAS NLMIXED procedure (see Appendix 1 for example code), the authors developed a Fortran program to maximize (4) and to compute the observed information matrix and its inverse. This program is build around the NAG library subroutines D01BBF and D01FBF for the numerical integration and E04UCF for the optimization (NAG, 1993). In some cases, besides $\boldsymbol{\xi}$ also the person specific effects $\theta_p$, $\eta_p$ and $\lambda_p$ are of special interest. Estimates of these parameters can be obtained from an empirical Bayes analysis of the postulated model.

## 3 Simulation study

In this section we discuss the results of a small simulation study. Four data sets, each containing responses of 2000 examinees on 40 items were generated. Sample 1 was generated under model (3)-(1) with extreme speededness ($\alpha = 2$ and $\beta = 2$). Sample 2 was generated under model (3)-(1) with moderately high speededness ($\alpha = 9$ and $\beta = 2$). Sample 3 was generated under model (3)-(1) with moderately low speededness ($\alpha = 20$ and $\beta = 2$). In this case, $\mathrm{E}(\eta_p) = 0.50$ for Sample 1, 0.82 for Sample 2 and 0.91 for Sample 3. Finally, a fourth sample was generated from a 1PL with random guessing. For the fourth sample, no speededness was generated. For all samples, the item difficulty parameters $\beta_i$, $i = 1, \ldots, I$, were assumed to be equal and fixed at the value -1. The complete list of parameter values is given in Table 1. The random effects are assumed to be independent. All computations were performed with the Fortran/NAG implementation. Computation times varied between 8 and 48 hours (on an Intel Pentium 3 M, 1.13 GHz, 512 MB of RAM).

The effect of test speededness is illustrated in Figure 3 (a), Figure 4 (a), Figure 5 (a) and Figure 6 (a), where we plot the empirical proportions correct answers (solid lines) together with the

Table 1: Parameter values for simulation study.

| Parameter | Sample 1 | Sample 2 | Sample 3 | Sample 4 |
|---|---|---|---|---|
| $\beta_1$ - $\beta_{40}$ | -1 | -1 | -1 | -1 |
| $c$ | 0.2 | 0.2 | 0.2 | 0.2 |
| $\sigma^2_\theta$ | 1 | 1 | 1 | 1 |
| $\alpha$ | 2 | 9 | 20 | - |
| $\beta$ | 2 | 2 | 2 | - |
| $\mu_\lambda$ | 0 | 0 | 0 | - |
| $\sigma^2_\lambda$ | 1 | 1 | 1 | - |

theoretical ones (broken lines), given by

$$\begin{aligned} \mathrm{E}(Y_{pi}) &= \mathrm{E}[\mathrm{E}(Y_{pi}|\theta_p, \eta_p, \lambda_p)] \\ &= \mathrm{E}(\pi_{pi}) \\ &= c + (1-c)\int_{\mathbb{R}} P_i(\theta_p)dG_1(\theta_p)\int_0^1\int_0^\infty \min\left\{1, \left[1-\left(\frac{i}{I}-\eta_p\right)\right]^{\lambda_p}\right\} dG_3(\lambda_p)dG_2(\eta_p), \end{aligned} \tag{5}$$

in case of (3), and by

$$\mathrm{E}(Y_{pi}) = c + (1-c)\int_{\mathbb{R}} P_i(\theta_p)dG_1(\theta_p),$$

in case of 1PL with random guessing, where $G_1$, $G_2$ and $G_3$ denote the distribution functions of $\theta_p$, $\eta_p$ and $\lambda_p$ respectively, versus item number. Since all $\beta_i$ are equal, these proportions should not depend on item number in the absence of test speededness (see Figure 6 (a)). Clearly, test speededness decreases the probability of a correct answer for end-of-test items. Of course, the ultimate effect depends on the distribution of the speededness point and rate.

In Figure 3 (b) and (c) we illustrate the effect of test speededness on the $\beta_i$ estimates. For items early in the test, the $\beta_i$ estimates obtained with the test speededness model (solid line) and the 1PL with random guessing model (broken line) agree quite well. However, from a certain point on the $\beta_i$ estimates obtained from fitting a 1PL with guessing model diverge from those obtained under the speededness model. As is clear from the figures, ignoring test speededness causes upward biased estimates of the item difficulty estimates, a result that is consistent with the item response theory literature. In Figure 3 (d), (e) and (f) we show the theoretical density functions (solid lines) of the person ability, speededness point and speededness rate respectively, together with the fitted densities (dotted lines) and approximate 95% confidence intervals for the true density function (broken-dotted lines). These approximate confidence intervals were obtained by applying the delta method; for more details we refer to Appendix 3. In Figure 4, Figure 5 and Figure 6 we present the corresponding estimation results for Sample 2, Sample

3 and Sample 4 respectively. For the data generated from the 1PL with guessing model, the observed information matrix of model (3)-(1) was ill-conditioned and could not be inverted. As a consequence, we could not construct confidence intervals for the random effect density functions.

The fit of the speededness model can be further evaluated by comparing the estimated theoretical proportions correct answers, obtained by plugging the maximum likelihood estimates for the model parameters into (5), with the corresponding empirical proportions. These estimated theoretical proportions are drawn by broken-dotted lines in Figure 3 (a), Figure 4 (a), Figure 5 (a) and Figure 6 (a). Clearly, the estimated and empirical proportions correct answers are almost indistinguishable, indicating a very good fit of the model.

In Table 2 we compare model (3) with the 1PL model with random guessing in terms of $-2\log L$, the Akaike information criterion (AIC) and the Schwarz Bayes information criterion (BIC). The 1PL model with random guessing is nested in the test speededness model and hence its values for $-2\log L$ will always be larger than the ones for model (3). The difference in $-2\log L$ values can be used to construct a likelihood ratio test for the 1PL model with random guessing. As the hypothesis of interest is on the boundary of the parameter space, one has to adjust the distribution of the likelihood ratio test under the null hypothesis. In general, the asymptotic null distribution for the likelihood ratio test statistic for testing a null hypothesis which allows for $J$ correlated random effects versus an alternative of $J+1$ correlated random effects, is a mixture of a $\chi^2_J$ and a $\chi^2_{J+1}$, with equal probability 0.5. For the more general case where one compares models with $J$ and $J+q$ ($q>1$) correlated random effects, the null distribution is a mixture of $\chi^2$ random variables, the weights of which can only be calculated analytically in a number of special cases (Raubertas *et al.*, 1986, Shapiro, 1988). Finally, in case of testing a model with $J$ independent random effects versus a model with $J+q$ ($q \geq 1$) independent random effects, as considered in this simulation, the asymptotic null distribution is a binomial mixture of $\chi^2$ random variables, see Verbeke and Molenberghs (2003). Inference concerning the fixed effects can be drawn on the basis of a classical likelihood ratio test or using a Wald test statistic. For all cases considered, AIC selects the appropriate model, i.e. the test speededness model for Sample 1, Sample 2 and Sample 3 and the 1PL with random guessing model for Sample 4. The BIC penalizes complex models more heavily than the AIC and hence indicates also for Sample 3 the 1PL model with random guessing as the most appropriate one.

Table 2: Goodness-of-fit of test speededness model versus 1PL with guessing.

| | Sample 1 | | Sample 2 | | Sample 3 | | Sample 4 | |
|---|---|---|---|---|---|---|---|---|
| | speeded | 1PL guessing | speeded | 1PL guessing | speeded | 1PL guessing | speeded | 1PL guessing |
| $-2\log L$ | 93204 | 93882 | 86220 | 86379 | 84026 | 84035 | 82804 | 82808 |
| AIC | 93296 | 93966 | 86312 | 86463 | 84118 | 84119 | 82896 | 82892 |
| BIC | 93554 | 94201 | 86570 | 86698 | 84375 | 84354 | 83153 | 83127 |

Finally, we mention that, except for $\beta$, accurate estimation of the parameters related to the distributions of the speededness random effects $\eta$ and $\lambda$ is getting more difficult if test speededness comes in late. This is illustrated in Table 3 where we show the standard errors of all

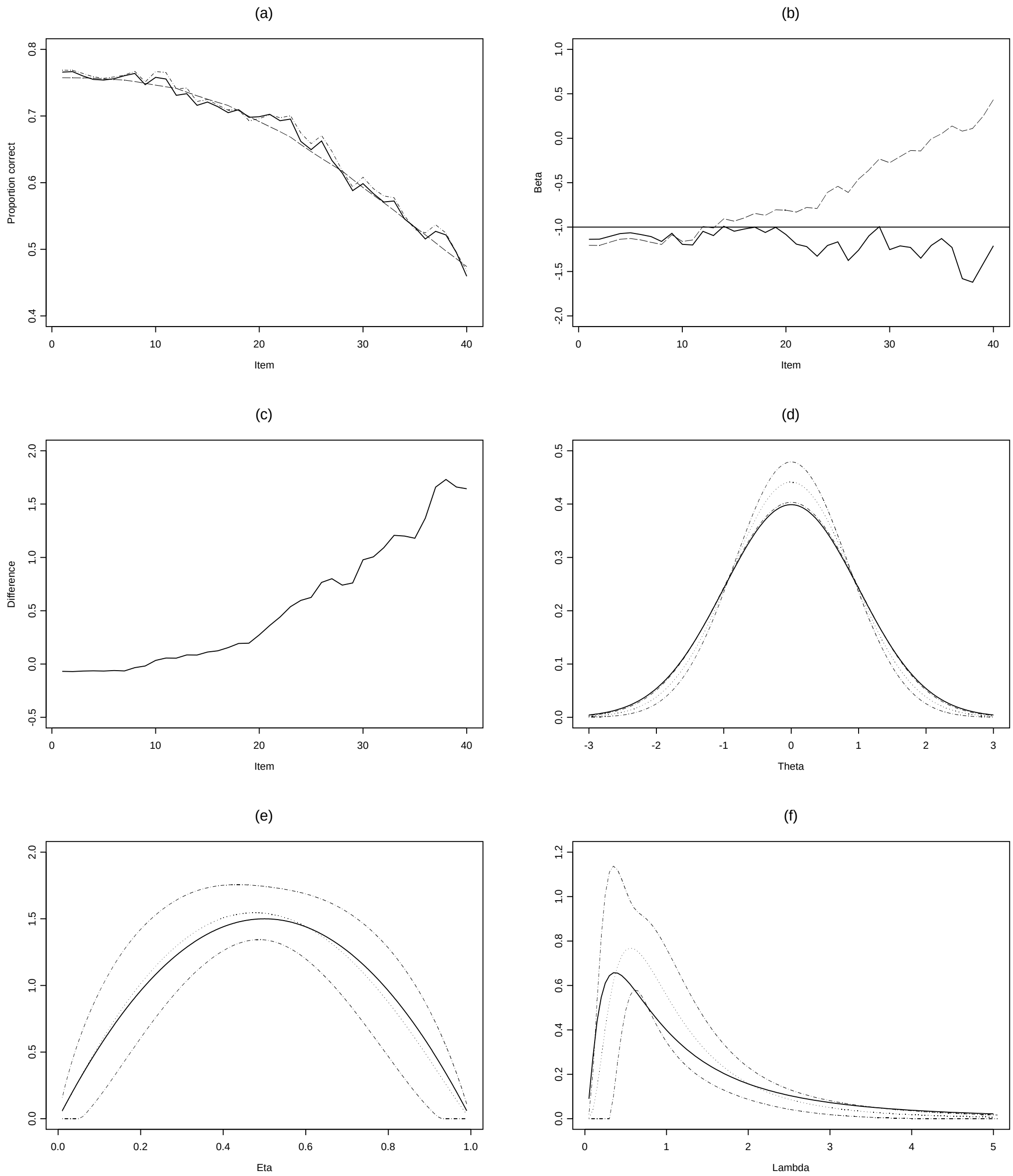


Figure 3: Results for Sample 1 (a) proportion correct versus item number: empirical (solid line), theoretical with true parameter values (broken line), theoretical with estimated parameter values (broken-dotted line), (b) estimated item difficulty parameters under (3)-(1) (solid line) and 1PL with guessing (broken line), (c) difference between item difficulty estimates, (d) distribution of $\theta$: theoretical (solid line) and fitted (dotted line), (e) distribution of $\eta$: theoretical (solid line) and fitted (dotted line) and (f) distribution of $\lambda$: theoretical (solid line) and fitted (dotted line). In (d), (e) and (f) the broken-dotted lines are 95% confidence intervals.

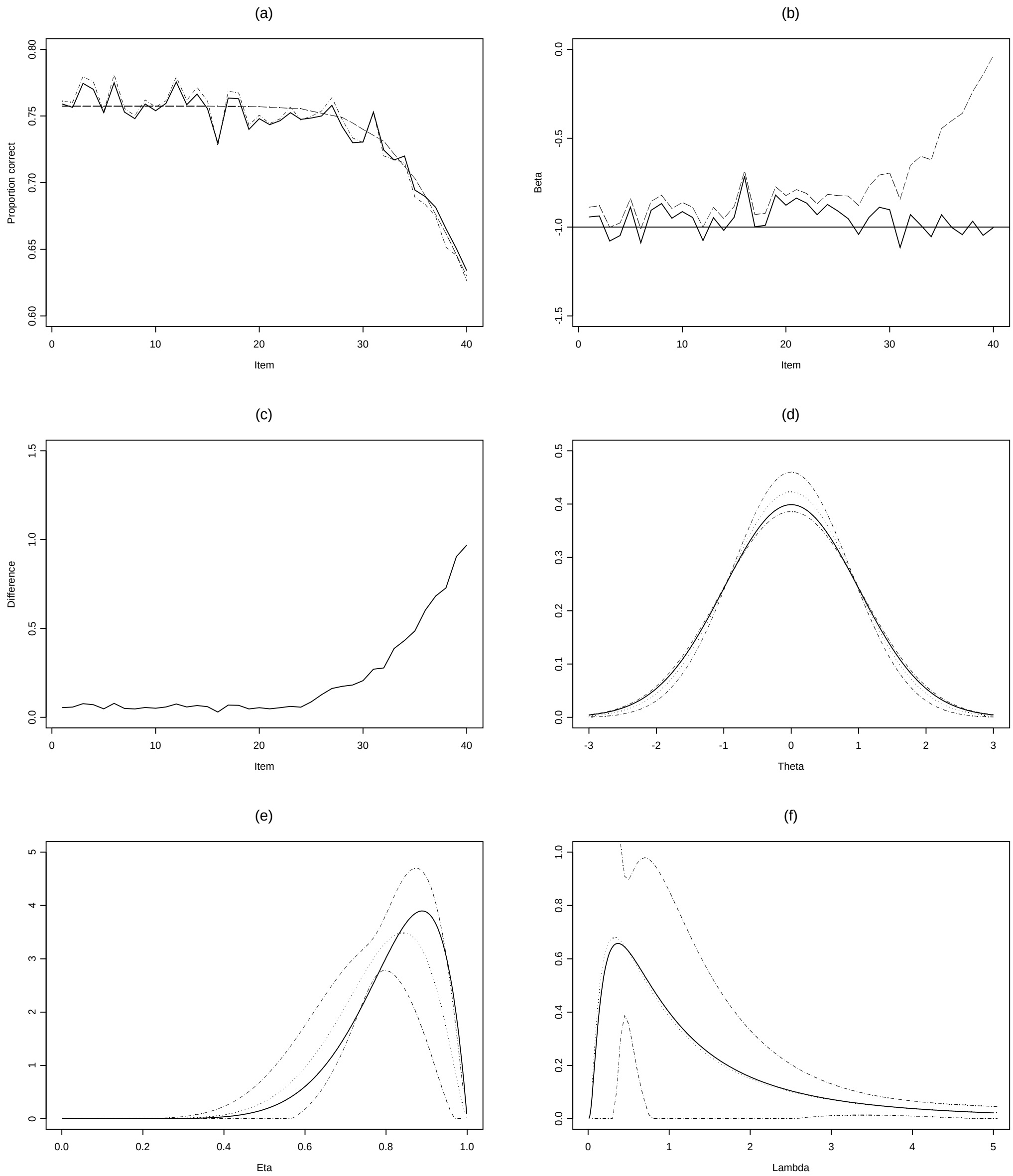


Figure 4: Results for Sample 2 (a) proportion correct versus item number: empirical (solid line), theoretical with true parameter values (broken line), theoretical with estimated parameter values (broken-dotted line), (b) estimated item difficulty parameters under (3)-(1) (solid line) and 1PL with guessing (broken line), (c) difference between item difficulty estimates, (d) distribution of $\theta$: theoretical (solid line) and fitted (dotted line), (e) distribution of $\eta$: theoretical (solid line) and fitted (dotted line) and (f) distribution of $\lambda$: theoretical (solid line) and fitted (dotted line). In (d), (e) and (f) the broken-dotted lines are 95% confidence intervals.

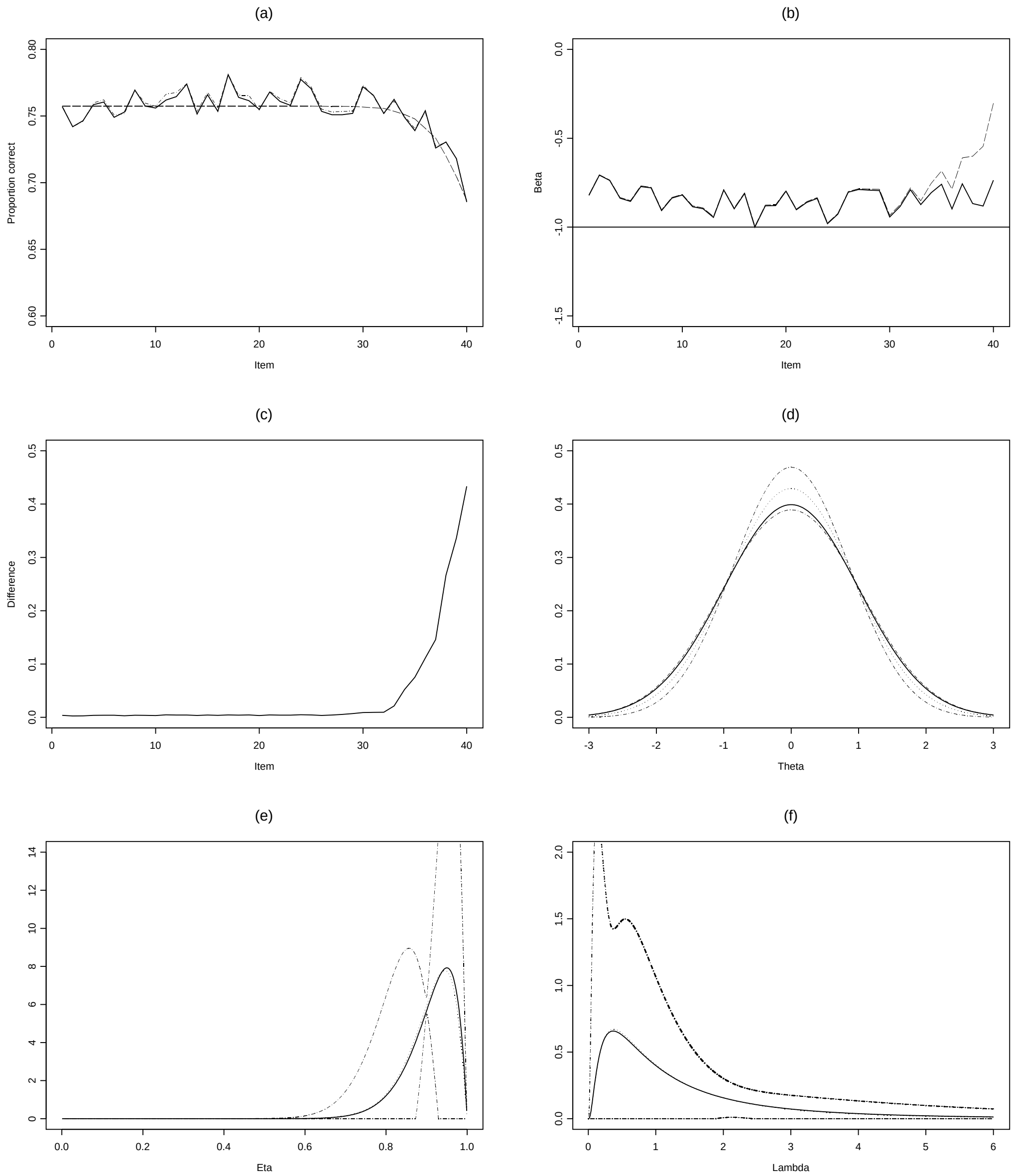


Figure 5: Results for Sample 3 (a) proportion correct versus item number: empirical (solid line), theoretical with true parameter values (broken line), theoretical with estimated parameter values (broken-dotted line), (b) estimated item difficulty parameters under (3)-(1) (solid line) and 1PL with guessing (broken line), (c) difference between item difficulty estimates, (d) distribution of $\theta$: theoretical (solid line) and fitted (dotted line), (e) distribution of $\eta$: theoretical (solid line) and fitted (dotted line) and (f) distribution of $\lambda$: theoretical (solid line) and fitted (dotted line). In (d), (e) and (f) the broken-dotted lines are 95% confidence intervals.

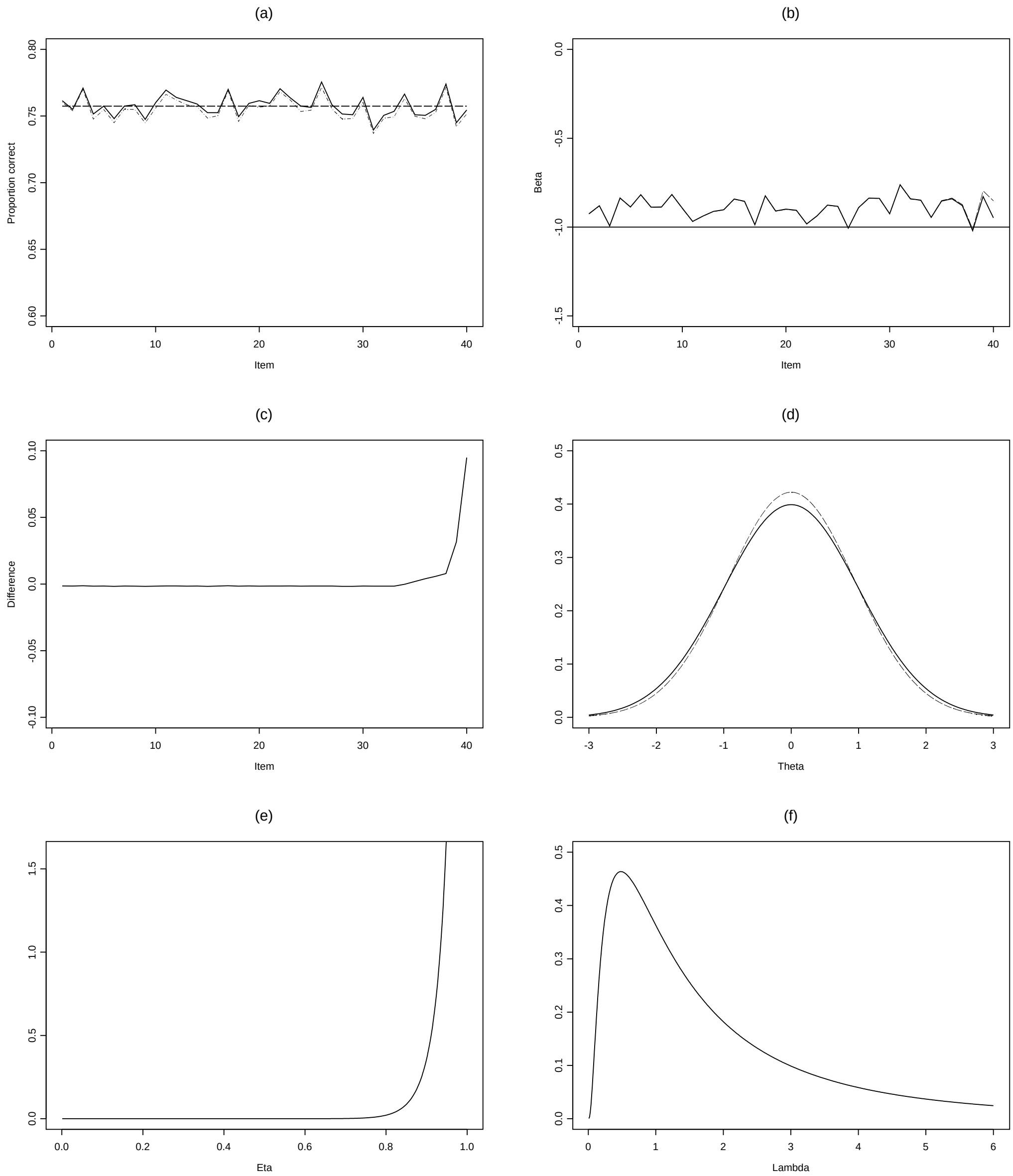


Figure 6: Results for Sample 4 (a) proportion correct versus item number: empirical (solid line), theoretical with true parameter values (broken line), theoretical with estimated parameter values (broken-dotted line), (b) estimated item difficulty parameters under (3)-(1) (solid line) and 1PL with guessing (broken line), (c) difference between item difficulty estimates, (d) distribution of $\theta$: theoretical (solid line) and fitted (broken line), (e) fitted density function of $\eta$ and (f) fitted density function of $\lambda$.

parameter estimates for random effect distribution parameters. As is clear, the standard errors of $\hat{\alpha}$, $\hat{\mu}_\lambda$ and $\hat{\sigma}^2_\lambda$ all increase from Sample 1 to Sample 2 and Sample 3. Caution is in order when comparing the standard errors of $\hat{\alpha}$ across samples, as this parameter increases from 2 in Sample 1 up to 20 in Sample 3. For this reason we also included the coefficient of variation (CV) of this parameter in the final row of Table 3. The other parameters were fixed in the simulation design and hence their standard errors can be safely compared across samples. Note that the standard errors of the $\beta$ estimates are quite stable over the samples. This is not completely unexpected as the $\beta$ parameter governs the shape of the right tail of the Beta distribution and hence, unlike the $\alpha$ estimates, is less sensitive to the point where test speededness starts.

Table 3: Standard errors of parameter estimates related to random effect distributions.

| Parameter | Sample 1 | Sample 2 | Sample 3 |
|---|---|---|---|
| $\hat{\sigma}^2_\theta$ | 0.040 | 0.044 | 0.046 |
| $\hat{\alpha}$ | 0.156 | 1.037 | 10.679 |
| $\hat{\beta}$ | 0.177 | 0.171 | 0.178 |
| $\hat{\mu}_\lambda$ | 0.093 | 0.278 | 0.318 |
| $\hat{\sigma}^2_\lambda$ | 0.105 | 0.761 | 0.911 |
| $CV(\hat{\alpha})$ | 0.078 | 0.115 | 0.534 |

# 4 Application to mathematics placement test

Data from Form 1 of the 2004 administration of a mathematics placement test at a large, selective Midwestern university were analyzed for test speededness using model (3)-(1). The data set contains response profiles of 3447 students. The mathematics placement test included 75 operational and 10 pilot items covering mathematics basics, college algebra and trigonometry and is designed to be completed in 90 minutes. All items had 5 alternatives. Because the item-total correlations for the last five pilot items (locations 45, 55, 65, 75 and 85) were poor, these items were dropped, resulting in an analysis of 80 items.

In Table 4, we compare the test speededness model and the 1PL model with guessing in terms of $-2 \log L$, AIC and BIC. As is clear, all criteria indicate the test speededness model as the most appropriate one to describe these data. We now further evaluate the fit of the proposed test speededness model. In Figure 8 (a) we plot the empirical (solid line) and estimated theoretical (broken line) proportions correct answers versus the item number. The proportions correct answers clearly tend to decrease when considered as a function of item number. This does not necessarily indicate test speededness as the items may simply be ordered according to item difficulty, with the more difficult items near the end of the test. Note however that the test speededness model produces an almost perfect fit to the data: in Figure 8 (a) the estimated theoretical and empirical proportions correct answers are almost indistinguishable. This goodness-of-fit evaluation clearly only involves marginal probabilities and hence only gives a partial picture of the absolute model fit. To evaluate the absolute goodness-of-fit we have used a parametric bootstrap approach. In this we compare the empirical item characteristic curves (ICCs) with those obtained from repeated sampling from the proposed test speededness model with parameters replaced by their maximum likelihood estimates. If the model really fits the data, the observed ICCs should be in line with the simulated ones. The bootstrap procedure was implemented with a uniform (-4,4) distribution for the person ability parameters. This choice was made in order to obtain also reliable estimates of the ICCs in the lower and upper ranges of ability. In Figure 7 we show for some items the empirical ICCs (solid lines) together with those obtained from 100 bootstrap iterations (dots). As is clear from this plot, all empirical ICCs are contained in the confidence band based on the bootstrap samples, giving further evidence in favour of the model fit. The bootstrap goodness-of-fit results for the other items are similar to those given in Figure 7.

Table 4: Mathematics placement test data: Goodness-of-fit of test speededness model versus 1PL with guessing.

| | speeded | 1PL guessing |
|---|---|---|
| $-2\log L$ | 256018 | 256954 |
| AIC | 256190 | 257118 |
| BIC | 256719 | 257622 |

Further estimation results are graphically represented in Figure 8. In Figure 8 (b) and (c) we compare the estimates for the item difficulty parameters obtained under the test speededness model (3) with those obtained under 1PL with random guessing. The $\beta_i$ under (3) are in the range $[-3.95; 2.68]$. The estimated difficulties of end-of-test items are clearly larger under the 1PL with random guessing model than under the test speededness model. Moreover, the difference between the two item difficulty estimates tends to increase in item number, see Figure 8 (c). In Figure 8 (d), (e) and (f) we plot the fitted random effect density functions (solid lines) together with 95% confidence intervals (broken-dotted lines). For the speededness point parameter $\eta_p$, we obtained $\hat{\alpha} = 1.312$ and $\hat{\beta} = 0.767$, yielding a Beta distribution with mean 0.631 and variance 0.076. Concerning the speededness rate $\lambda_p$, the estimates are $\hat{\mu}_\lambda = -1.127$ and $\hat{\sigma}_\lambda = 1.140$, resulting in a log-normal distribution with mean 0.621 and variance 1.029.

# 5 Discussion and conclusion

In this paper we proposed an item response theory model dealing with test speededness. The model can be seen as consisting of two random processes, a Rasch process and a random guessing process, with the random guessing process gradually taking over from the Rasch process. Both change point and change rate are considered as random effects in order to model examinee differences in both respects. The model improves on the hybrid model of Yamamoto and

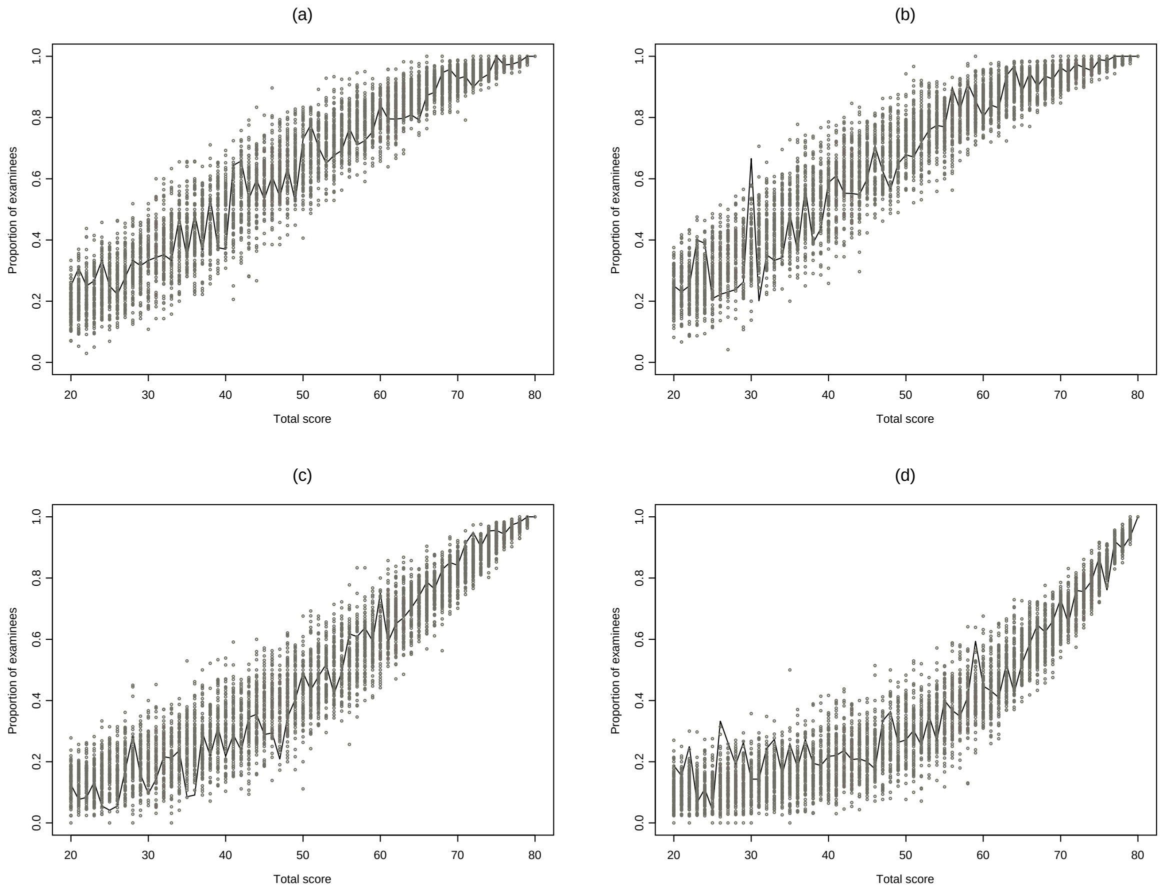


Figure 7: Mathematics placement test data: empirical ICCs of the real data (solid line) plotted against bootstrap results: (a) item 16, (b) item 34, (c) item 60 and (d) item 79.

Everson (1997) in the sense that examinees do not switch immediately to random guessing once they become speeded. The model also extends the mixture Rasch model of Bolt *et al.* (2002), by allowing examinees to become speeded at different points in the test. From the simulation study we may conclude that recovery of the parameter values of the test speededness model is rather good and that the model can be differentiated from 1PL with guessing or 3PL by using information criteria such as AIC and BIC. Inference concerning the fixed effects of the proposed model can be drawn using Wald tests.

The model we presented is an instantiation of a more general category of models with gradual change in a series of repeated observations. For example, in a learning experiment one may start with guessing because one has no insight in how to solve the items, whereas later in the series a gradual shift may occur to a more appropriate strategy to actually solve the items, thanks to learning. This change process could be modeled in a way that is complementary to

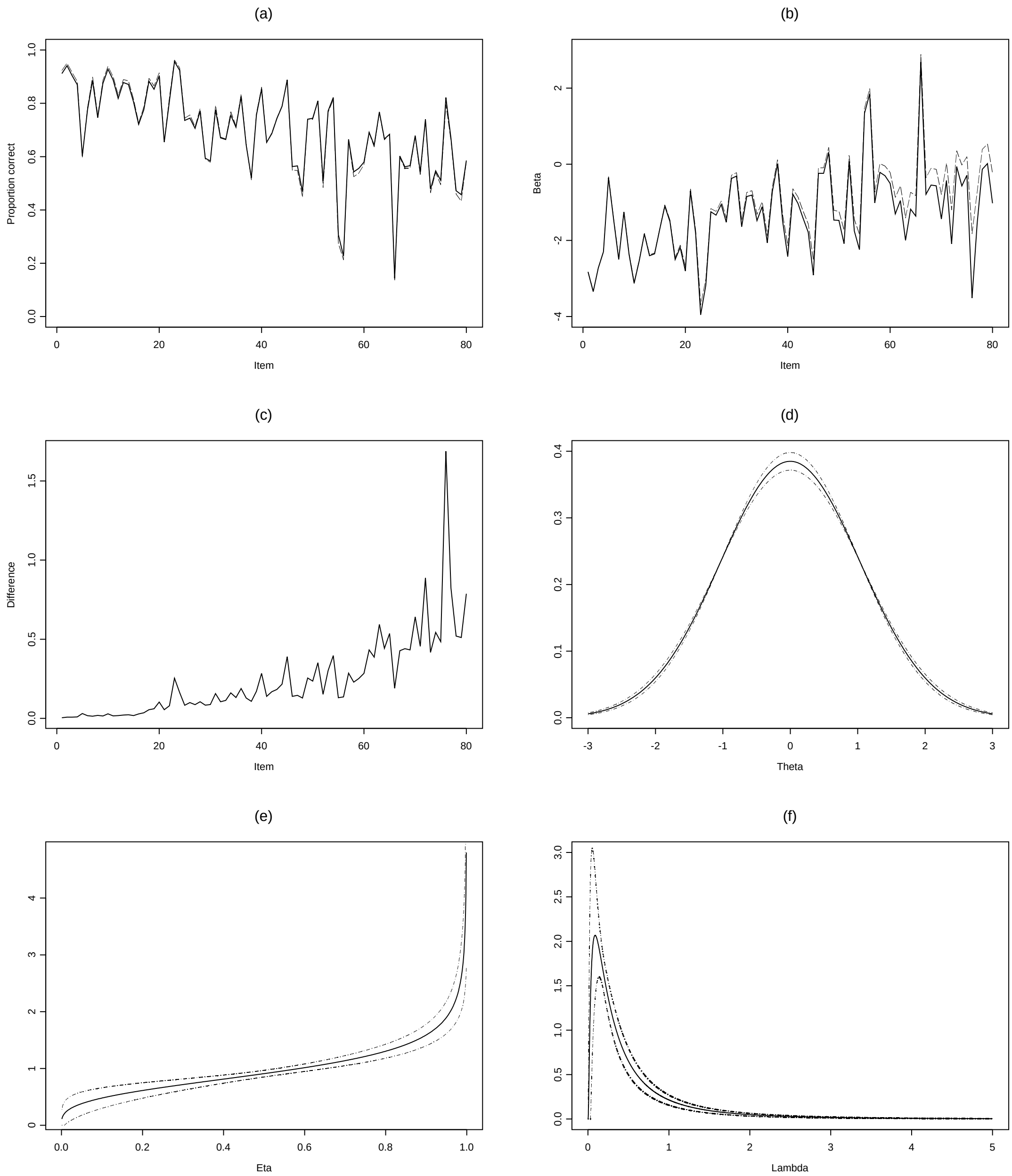


Figure 8: Mathematics placement test data: (a) proportion correct versus item number: empirical (solid line), theoretical with estimated parameter values (broken line), (b) estimated item difficulty parameters under (3)-(1) (solid line) and 1PL with guessing (broken line), (c) difference between item difficulty estimates, (d) fitted distribution of $\theta$, (e) fitted distribution of $\eta$ and (f) fitted distribution of $\lambda$. In (d), (e) and (f) the broken-dotted lines are 95% confidence intervals.

the speededness model, with a transition from guessing to solving instead of a transition from solving to guessing as for the case of speededness. From a general perspective, one may consider any transition between two strategies or two principles during a series of repeated observations, given that the two strategies or principles correspond to models that can also be estimated separately. This opens up a rather broad category of applications in psychology and in other disciplines.

The model considered assumes a dichotomous response (incorrect/correct) with test speededness gradually degrading responses towards incorrect answers. However, besides more frequent wrong answers, test speededness may also result in omitted answers. This omission may, next to test speededness, also depend on ability and hence dropout is, using the terminology of Little and Rubin (1987), missing not at random (MNAR). An early attempt to model dropout in test data can be found in Lord (1983), where a trinomial response model (omit/incorrect/correct) is proposed with dropout being examinee specific. Extensions of this model including test speededness are worthwhile considering. Next to this model, the selection and pattern-mixture models (see e.g. Glynn *et al.*, 1986), two popular dropout models in the biomedical sciences, may also deserve attention in this respect. This is a topic of ongoing research.

## Acknowledgements

The research reported in this paper was supported by IAP P5/24 and GOA/2005/04, both awarded to Paul De Boeck and Iven Van Mechelen.

## References

[1] Birnbaum, A., 1968. Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. and Novick, M.R. (Eds), *Statistical Theories of Mental Test Scores*, pp. 394-479. Addison-Wesly.

[2] Bolt, D.M., Cohen, A.S. and Wollack, J.A., 2002. Item parameter estimation under conditions of test speededness: application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, **39**, 331-348.

[3] De Boeck, P. and Wilson, M., 2004. *Explanatory Item Response Models - A Generalized Linear and Nonlinear Approach*. Springer.

[4] Douglas, J., Kim, H.R., Habing, B. and Gao, F., 1998. Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, **23**, 129-151.

[5] Ferrara, S., Huynh, H. and Michaels, H., 1999. Contextual explanations of local dependence in item clusters in a large-scale hands-on science performance assessment. *Journal of Educational Measurement*, **36**, 119-140.

[6] Glynn, R.J., Laird, N.M. and Rubin, D.B., 1986. Selection modelling versus mixture modelling with nonignorable nonresponse. In Wainer, H. (Ed), *Drawing Inferences from Self-Selected Samples*, pp 115-142. Springer

[7] Hutchinson, T.P., 1991. *Ability, Partial Information, Guessing: Statistical Modelling Applied to Multiple-Choice Tests*. Rumsby Scientific Publishing.

[8] Joe, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman & Hall.

[9] Lee, G., Kolen, M.J., Frisbie, D.A. and Ankenmann, R.D., 2001. Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, **25**, 357-372.

[10] Lehmann, E.L. and Casella, G., 2003. *Theory of Point Estimation*. Springer.

[11] Little, R.J.A. and Rubin, D.B., 1987. *Statistical Analysis with Missing Data*. Wiley.

[12] Lord, F.M., 1983. Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika*, **48**, 477-482.

[13] NAG, (1993). *NAG Fortran Library Manual - Mark 19*. The Numerical Algorithms Group Limited.

[14] Nelsen, R.B., 1999. *An Introduction to Copulas*. Springer.

[15] Oshima, T.C., 1994. The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement*, **31**, 200-219.

[16] Rasch, G., 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen, Denmark.

[17] Raubertas, R.F., Lee, C.I.C. and Nordheim, E.V., 1986. Hypothesis tests for normal means constrained by linear inequalities. *Communications in Statistics - Theory and Methods*, **15**, 2809-2833.

[18] Rost, J., 1990. Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement*, **14**, 271-282.

[19] Shapiro, A., 1988. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, **56**, 49-62.

[20] Sireci, S.G., Thissen, D. and Wainer, H., 1991. On the reliability of testlet-based tests. *Journal of Educational Measurement*, **28**, 237-247.

[21] Sklar, A., 1959. Fonctions de repartition a $n$ dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229-231.

[22] Tanner, M.A., 1996. *Tools for Statistical Inference - Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer.

[23] Thissen, D., Steinberg, L. and Mooney, J., 1989. Trace lines for testlets: a use of multiple-categorical response models. *Journal of Educational Measurement*, **26**, 247-260.

[24] Tuerlinckx, F. and De Boeck, P., 2001. The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods*, **6**, 181-195.

[25] Verbeke, G. and Molenberghs, G., 2003. The use of score tests for inference on variance components. *Biometrics*, **59**, 254-262.

[26] Wainer, H. and Thissen, D., 1996. How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, **15**, 22-29.

[27] Wollack, J.A. and Cohen, A.S., 2004. A model for simulating speeded test data. Technical report.

[28] Yamamoto, K., 1987. A model that combines IRT and latent class models. PhD thesis, University of Illinoiis, Champaign-Urbana.

[29] Yamamoto, K. and Everson, H., 1997. Modeling the effects of test length and test time on parameter estimation using the hybrid model. In Rost, J. and Langeheine, R. (Eds.), *Applications of Latent Trait and Latent Class Models in the Social Sciences*, pp. 89-99. Waxmann, New York.

[30] Yen, W.M., 1984. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, **8**, 125-145.

[31] Yen, W.M., 1993. Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement*, **30**, 187-213.

## Appendix 1: Example SAS code

```
data simdata1;
infile 'c:\irm\speeded\paper\simdata1.txt';
input y
nr person x1-x40;
nr_n=nr/40;
run;


proc nlmixed data=simdata1 method=gauss noad technique=newrap
maxiter=500 maxfu=5000 qpoints=5;

parms b1-b40=-1 c=.2 s2t=1 a=9 b=2 ml=0 s2l=1 ;

beta =
b1*x1+b2*x2+b3*x3+b4*x4+b5*x5+b6*x6+b7*x7+b8*x8+b9*x9+b10*x10+
b11*x11+b12*x12+b13*x13+b14*x14+b15*x15+b16*x16+b17*x17+b18*x18+b19*x19+b20*x20+
b21*x21+b22*x22+b23*x23+b24*x24+b25*x25+b26*x26+b27*x27+b28*x28+b29*x29+b30*x30+
b31*x31+b32*x32+b33*x33+b34*x34+b35*x35+b36*x36+b37*x37+b38*x38+b39*x39+b40*x40;

eta=betainv(probnorm(et),a,b);

lambda=exp(la);

r=exp(theta-beta)/(1+exp(theta-beta));

s=(1-(nr_n-eta))**lambda;

if (s >=1) then pr=c+(1-c)*r;

else pr=c+(1-c)*r*s;

model y ~ binary(pr);

random theta la et ~ normal([0,ml,0],[s2t,0,s2l,0,0,1])
subject=person;

run;
```

## Appendix 2

**Definition 1** *A $n$-copula is a function $C : [0,1]^n \to [0,1]$ with the following properties*

1. *for every $\boldsymbol{u} \in [0,1]^n$ with at least one coordinate equal to 0, $C(\boldsymbol{u}) = 0$,*
2. *if all coordinates of $\boldsymbol{u}$ are 1 except $u_k$ then $C(\boldsymbol{u}) = u_k$,*
3. *for all $\boldsymbol{a}, \boldsymbol{b} \in [0,1]^n$ with $\boldsymbol{a} \leq \boldsymbol{b}$ the volume of the hyperrectangle with corners $\boldsymbol{a}$ and $\boldsymbol{b}$ is positive, i.e.*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C(u_{i_1}, \ldots, u_{i_n}) \geq 0$$

*where $u_{i_1} = a_i$ and $u_{i_2} = b_i$.*

So essentially a $n$-copula is a $n$-dimensional distribution function on $[0,1]^n$ with standard uniform marginal distributions. The next theorem, due to Sklar, is central to the theory of copulas and forms the basis of the applications of that theory to statistics.

**Theorem 1 Sklar (1959)** *Let $\boldsymbol{X}' = (X_1, \ldots, X_n)$ be a random vector with joint distribution function $F_{\boldsymbol{X}}$ and marginal distribution functions $F_i$, $i = 1, \ldots, n$. Then there exists a copula $C$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_n(x_n)). \tag{6}$$

*If $F_1, \ldots, F_n$ are all continuous then $C$ is unique, otherwise $C$ is uniquely determined on Ran $F_1 \times \cdots \times$ Ran $F_n$. Conversely, given a copula $C$ and marginal distribution functions $F_1, \ldots, F_n$, the function $F_{\boldsymbol{X}}$ as defined by (6) is a joint distribution function with margins $F_1, \ldots, F_n$.*

As is clear, Sklar's theorem separates a joint distribution into a part that describes the dependence structure (the copula) and parts that describe the marginal behavior (the marginal distributions). For further details on copula functions we refer to Joe (1997) and Nelsen (1999).

**Proposition 1** *Consider a $n$-dimensional random vector $\boldsymbol{X}$ with joint distribution function $G$ and continuous marginal distribution functions $G_1, \ldots, G_n$. Assume that $G$ is characterized by a normal dependence function (copula) $C$ i.e.*

$$G(x_1, \ldots, x_n) = C(G_1(x_1), \ldots, G_n(x_n))$$

*with*

$$C(u_1, \ldots, u_n) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_n)} \frac{1}{(2\pi)^{n/2} |\boldsymbol{R}|^{1/2}} \, e^{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{R}^{-1}\boldsymbol{z}} d\boldsymbol{z}$$

*in which $\boldsymbol{R}$ denotes a (positive definite) correlation matrix and $\Phi^{-1}$ is the inverse standard normal distribution function. Then the random variables*

$$Y_i = \Phi^{-1}(G_i(X_i)), \quad i = 1, \ldots, n,$$

*are jointly distributed as multivariate normal.*

**Proof:** Denote the joint distribution function of $Y_1, \ldots, Y_n$ by $H$. Then

$$\begin{aligned} H(y_1, \ldots, y_n) &= P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) \\ &= P(\Phi^{-1}(G_1(X_1)) \leq y_1, \ldots, \Phi^{-1}(G_n(X_n)) \leq y_n) \\ &= P(X_1 \leq G_1^{-1}(\Phi(y_1)), \ldots, X_n \leq G_n^{-1}(\Phi(y_n))) \\ &= C(\Phi(y_1), \ldots, \Phi(y_n)) \\ &= \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} \frac{1}{(2\pi)^{n/2}|\boldsymbol{R}|^{1/2}} \, e^{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{R}^{-1}\boldsymbol{z}} d\boldsymbol{z}, \end{aligned}$$

which is the distribution function of a multivariate normal distribution. ∎

# Appendix 3

In this appendix we elaborate on the construction of approximate $100(1-\alpha)\%$ confidence intervals for the random effect density functions. Let $g(x;\boldsymbol{\omega})$ denote a density function depending on an unknown parameter vector $\boldsymbol{\omega}$. A straightforward point estimator for $g$ at a point $x$ is obtained by replacing $\boldsymbol{\omega}$ by an estimator $\hat{\boldsymbol{\omega}}$. To construct an approximate confidence interval for $g(x;\boldsymbol{\omega})$ we need the limiting distribution of $\sqrt{n}(g(x;\hat{\boldsymbol{\omega}}) - g(x;\boldsymbol{\omega}))$. The following general result, often referred to as the delta method, provides the limiting distribution of functions of limiting normal random vectors (see e.g. Lehmann and Casella, (2003) and Tanner, (1996)).

**Theorem 2** *Let $\hat{\boldsymbol{\theta}}_n$ denote an estimator of the p-dimensional vector $\boldsymbol{\theta}$ satisfying*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \Sigma),$$

*as $n \to \infty$. Suppose $g$ a real valued function of $\boldsymbol{\theta}$, defined on and continuously differentiable in a neighborhood $\delta$ of the parameter point $\boldsymbol{\theta}$ and set $\boldsymbol{\kappa}(\boldsymbol{\theta}) = \partial g(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Then, for $n \to \infty$,*

$$\sqrt{n}(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta})) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\kappa}(\boldsymbol{\theta})'\Sigma\boldsymbol{\kappa}(\boldsymbol{\theta})),$$

*provided $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is not equal to zero.*

**Corollary 1** *Consider a density (or probability) function $g(x;\boldsymbol{\omega})$ depending on a parameter vector $\boldsymbol{\omega}$ and assume that $g$ satisfies all conditions of Theorem 2. Let $\hat{\boldsymbol{\omega}}$ denote the maximum likelihood estimator for $\boldsymbol{\omega}$ based on a sample $X_1, \ldots, X_n$ of independent and indentically distributed random variables from $g$ and suppose that all regularity conditions for asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ are satisfied. Then, denoting by $I(\hat{\boldsymbol{\omega}})$ the observed Fisher information matrix of $\boldsymbol{\omega}$, i.e.*

$$I(\hat{\boldsymbol{\omega}}) = -\frac{1}{n}\sum_{i=1}^{n} \left.\frac{\partial^2 \ln g(X_i;\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}'}\right|_{\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}},$$

*the interval*

$$\left[g(x;\hat{\boldsymbol{\omega}}) - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\boldsymbol{\kappa}(\hat{\boldsymbol{\omega}})'I^{-1}(\hat{\boldsymbol{\omega}})\boldsymbol{\kappa}(\hat{\boldsymbol{\omega}})}{n}};\; g(x;\hat{\boldsymbol{\omega}}) + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\boldsymbol{\kappa}(\hat{\boldsymbol{\omega}})'I^{-1}(\hat{\boldsymbol{\omega}})\boldsymbol{\kappa}(\hat{\boldsymbol{\omega}})}{n}}\right],$$

*with $\Phi^{-1}$ denoting the standard normal quantile function, is an approximate $100(1-\alpha)\%$ confidence interval for $g(x;\boldsymbol{\omega})$.*

In the following subsections we provide the elements to compute confidence intervals for the random effects density functions. The elements of the observed information matrix and the associated Fortran code can be obtained on request from the first author.

### Density function of person ability

Concerning the examinee ability $\theta_p$, we assumed a normal density function. To be specific

$$g_1(\theta_p;\sigma_\theta^2) = \frac{\exp[-\theta_p^2/(2\sigma_\theta^2)]}{\sqrt{2\pi}\sigma_\theta},$$

and hence

$$\kappa = \frac{\partial g_1(\theta_p;\sigma_\theta^2)}{\partial\sigma_\theta^2} = \frac{g_1(\theta_p;\sigma_\theta^2)}{2\sigma_\theta^2}\left(\frac{\theta_p^2}{\sigma_\theta^2}-1\right).$$

### Density function of speededness point

The change point $\eta_p$ was assumed to be Beta distributed:

$$g_2(\eta_p;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\eta_p^{\alpha-1}(1-\eta_p)^{\beta-1}.$$

Let $\kappa_1 = \partial g_2(\eta_p;\alpha,\beta)/\partial\alpha$ and $\kappa_2 = \partial g_2(\eta_p;\alpha,\beta)/\partial\beta$. Then

$$\begin{aligned}
\kappa_1 &= g_2(\eta_p;\alpha,\beta)[\psi(\alpha+\beta)-\psi(\alpha)+\ln\eta_p],\\
\kappa_2 &= g_2(\eta_p;\alpha,\beta)[\psi(\alpha+\beta)-\psi(\beta)+\ln(1-\eta_p)],
\end{aligned}$$

where $\psi$ denotes the digamma function, i.e. $\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$.

### Density function of speededness rate

For the decay rate $\lambda_p$ we proposed a lognormal distribution:

$$g_3(\lambda_p;\mu_\lambda,\sigma_\lambda^2) = \frac{\exp[-(\ln\lambda_p-\mu_\lambda)^2/(2\sigma_\lambda^2)]}{\sqrt{2\pi}\sigma_\lambda\lambda_p}.$$

Let $\kappa_1 = \partial g_3(\lambda_p;\mu_\lambda,\sigma_\lambda^2)/\partial\mu_\lambda$ and $\kappa_2 = \partial g_3(\lambda_p;\mu_\lambda,\sigma_\lambda^2)/\partial\sigma_\lambda^2$. Then

$$\begin{aligned}
\kappa_1 &= g_3(\lambda_p;\mu_\lambda,\sigma_\lambda^2)\frac{\ln\lambda_p-\mu_\lambda}{\sigma_\lambda^2},\\
\kappa_2 &= \frac{g_3(\lambda_p;\mu_\lambda,\sigma_\lambda^2)}{2\sigma_\lambda^2}\left[\frac{(\ln\lambda_p-\mu_\lambda)^2}{\sigma_\lambda^2}-1\right].
\end{aligned}$$