Measurement and Research Department Reports 2000-1

A DOUBLE HAZARD MODEL FOR MENTAL SPEED

H.H.F.M. Verstralen
N.D. Verhelst
T.M. Bechger

Cito

Measurement and Research Department Reports 2000-1

A DOUBLE HAZARD MODEL FOR MENTAL SPEED

H.H.F.M. Verstralen
N.D. Verhelst
T.M. Bechger

Cito
Arnhem, 2000

A DOUBLE HAZARD MODEL FOR MENTAL SPEED

H.H.F.M. VERSTRALEN, N.D. VERHELST, and T.M. BECHGER
National Institute for Educational Measurement
Cito, Arnhem
the Netherlands

24.01.2000

Abstract

The administration of tests via the computer allows the registration of response times along with the actual response. This paper describes a model that combines these two kinds of data to estimate a subject latent variable usually called *mental speed*, but more appropriately called *mental power*. The model implies that the expected item score increases with invested time. Nevertheless, it allows for a decreasing expected item score with response time, which is sometimes found in experiments. This paradox is obtained by assuming that a subject not only stops working on a problem because of time pressure, but also when he has solved the problem. The model builds on a familiar framework of IRT models. An MML estimation procedure is developed, and model fit on the item level is evaluated using Lagrange multiplier tests.

1 Introduction

The administration of tests via the computer allows the registration of response times along with the actual response. This paper describes a model that combines these two kinds of data to estimate a subject latent variable known in the literature as *mental speed*, but more appropriately called *mental power*, because in the present context it is confronted with an item property that can be understood as resistance to solution. The item score reflects the insight the subject has gained about the presented problem within the observed response time. Insight is represented by a latent variable called *precision* that determines the item score distribution. Higher precision implies a higher expected item score. In this paper precision is not constant, but in general increases with response time. Mental power is defined as the increase of precision that a subject realizes per unit of time. As a result of greater the mental power a greater solution speed is realized.

There are few places in the literature where the simultaneous analysis of speed and accuracy is discussed. Even the psychometric literature primarily focuses on response time distributions, without regard to the response quality (Scheiblechner, 1979, 1985; Van Breukelen, 1995,1997; Maris, 1993). An exception is Verhelst, Verstralen, and Jansen (1997), who present an IRT model for mental power where the item scores and the total response time to the test are analyzed. The presumed absence of response times per item forced some less appealing restrictions on the model. Firstly, it had to be assumed that the time distribution is equal for all items. Consequently, the model looses credibility in situations were items are not of comparable difficulty or length. Secondly, the response times per item were assumed to be gamma distributed. This assumption was introduced for reasons of mathematical elegance, and not justified on substantial grounds. On the other hand, Storms and Delbeke (1992) demonstrate that the distributional assumptions on the response times are hardly relevant for other model characteristics. So, probably, even if this assumption would not pass empirical test, it will do little harm. Nevertheless, in the present study we assume that response time and quality are registered per item, and for each item a separate response time distribution is estimated.

The *conditional accuracy function* (CAF) is defined as the probability of a correct response as a function of response time. The models by Roskam, and Verhelst a.o. do not allow CAFs to be decreasing. On the other hand, Donders (1997) demonstrates that decreasing CAFs are a common phenomenon in practice. The present model was developed in order to accommodate both decreasing and increasing CAFs.

Traditionally, (e.g. Pachella, 1974) a distinction is made between macro trade-off and micro trade-off. Macro trade-off refers to the overall time pressure, and relates to the expected item score conditional on the mean reaction time. Formally,

$$\mathcal{E}(X|\mathcal{E}(T)), \tag{1}$$

with T the random variable response time. On the other hand, micro trade-off refers to the expected item score in relation to the decision to stop working on a particular item. Formally, micro trade-off can be represented as

$$\mathcal{E}(X|T). \tag{2}$$

All the models described by Pachella, based on feature detection theory, predict increasing precision with mean reaction time. However, the micro trade-off, that relates to the mean item score conditional on the reaction time of a particular subject may, under the Accumulator model, show a decreasing relation between precision and reaction time.

The model for micro trade-off in this report uses item response times and item scores to estimate mental power. It is rooted in the IRT tradition and, in that sense, a continuation of the work of Roskam, Van Breukelen, and Verhelst, a.o. The model is developed for polytomous items in general, and special attention is given to the CAFs.

It is a critical assumption in the present model that a subject can stop working on a problem for two reasons. He stops if he knows the correct answer to the item, or he stops if time pressure forces him to do so before a correct response is found. Therefore, the model is called the *double hazard model* (DHM). This assumption gives the model the property that, although we assume an increasing expected item score with invested time, a decreasing CAF may result. Not necessarily for all response times, however, because an item-person combination may show an increasing CAF for relatively fast responses, and a decreasing CAF for longer response times. This coincides with a common observation in educational evaluation. A student who recognizes the heart of the problem and quickly decides on an effective problem solving strategy has a high probability to solve the item within a relatively short time. His less lucky classmate, with equal mental power, may stumble on a less effective procedure, that not only takes longer, but is also likely to lead to an incorrect response. Nevertheless, also for this less lucky student it is not irrational or contradictory to state that a greater time investment in this problem increases his chances to better understand the problem and to produce a correct response.

2 General Properties of the Model

This section is about some general properties of the model with special attention to properties of the CAF. Consider a polytomous item with ordered categories $j = 0, ..., J$, where J will be referred to as the correct response. Let t denote the time (e.g. in seconds) that has passed from the moment a subject starts working on the item. We assume that the precision $\vartheta_v(t)$ of a subject v is increasing in t. This does not imply that the precision is unbounded. There may be an asymptotic value, so that it is not certain that the item will be solved, even with unlimited time. Let element $\pi_j(\vartheta_v(t))$ of the vector valued function $\underline{\pi}(\vartheta_v(t))$ ($\underline{\pi}= \pi_0, ..., \pi_J$) express the probability that a subject with precision $\vartheta_v(t)$ at time t scores in category j. The functional form of $\underline{\pi}$ and ϑ_v will not be specified until the next section. Here we discuss properties of the model that are independent of these specifications. Altough in some models $\vartheta(t)$ may be bounded from above, we can still in general discuss the reals as the domain of ϑ.

Denote with $f^{(n,\phi)}$ the nth derivative of f w.r.t. a parameter ϕ. In the sequel the following assumptions will be needed:

A-1 $\pi_J(\vartheta)$ is nondecreasing in ϑ, and $\lim_{\vartheta\to\infty} \pi_J(\vartheta) = 1$.

A-2 For $n = 1, 2$, $\pi_J^{(n,\vartheta)}(\vartheta)$ exists, and $\lim_{\vartheta\to\infty} \pi_J^{(n,\vartheta)}(\vartheta) = 0$.

A-3 For $n = 1, 2$, $\vartheta^{(n,\ln t)}(t)$ exists, and $\lim_{t\to\infty} \vartheta^{(n,\ln t)}(t)$ is bounded.

A-1 and A-2 are technical assumptions, and do not restrict any reasonable choice of model. Indeed they are satisfied by all models for polytomous item scores that we know. A-3, however, is more restrictive. It implies, for instance, that the model $\vartheta(t) = a + bt^c$ is unacceptable, since its first derivative w.r.t. $\ln t$ would have no limit for large t. But the model $\vartheta(t) = a + b \ln t$ is allowed. The characters a, b, and c denote arbitrary constants, with $b > 0$, and $c > 0$.

Denote with T_F the response time as a random variable from some distribution F. The distribution that represents the time pressure as perceived by the subjects is denoted as W, that is

$$W(t) = \Pr(T_W < t). \tag{3}$$

Formally, a subject starts by drawing a time t_W from W, works on the problem for at most t_W time units, and chooses a response at time t_W. But, if the subject has found the correct response before t_W, he will stop working on the problem and responds in the correct category J. If he fails to solve the problem before time t_W has elapsed he draws an incorrect response from the multinomial distribution

$$\Pr(X = j; t_W | X < J) = \frac{\pi_j(\vartheta(t_W))}{1 - \pi_J(\vartheta(t_W))}, \quad (\text{for } j \in \{0, ..., J-1\}). \tag{4}$$

The probability that subjects with precision function $\vartheta(t)$ respond correctly after time t is modeled by $\pi_J(\vartheta(t))$. If a subject responds correctly within time t, he found the correct response at t or before. That is, the conditional distribution of T given the correct response is given by

$$\Pr(T \leq t|X = J) = \pi_J(\vartheta(t)). \tag{5}$$

It follows that, formally, data generation in the DHM proceeds as follows:

1. Draw t_W from W
2. Draw t_π from π_J
3. If $t_\pi \leq t_W$ then the subject responds correctly at $t = t_\pi$
4. If $t_\pi > t_W$ then the subject draws his response from (4) at $t = t_W$.

We will now derive some general properties of the CAF under the DHM. First, the hazard function has to be introduced. Let $F(t)$ denote a time distribution with density function $f(t)$. The ratio

$$h_F(t) = \frac{f(t)}{1 - F(t)} \tag{6}$$

is defined for $F(t) < 1$ and is called a *hazard function.* The hazard function gives the conditional probability that some event occurs at $T = t$ given that it has not occurred yet. We adopt the convention that $f = F'$, where the prime denotes differentiation w.r.t. $\ln t$ rather than t, unless stated otherwise. In the model for W, and in the models for $\underline{\pi}$ and ϑ, introduced in the next section, differentiation w.r.t. $\ln t$ results in simpler formulas.

In the DHM there are two time distributions W and π_J, with corresponding hazards; $h_W(t)$ to stop at t as a result of time pressure, and $h_\pi(t)$ to stop because a correct response is found. Under the DHM the CAF is defined by

$$CAF(t) = \Pr(X = J|T = t) = \Pr(T_\pi \leq T_W|\min\{T_\pi, T_W\} = t) \tag{7}$$

By definition this conditional probability is given by

$$\begin{aligned} CAF(t) &= \frac{\pi'_J(1 - W)}{\pi'_J(1 - W) + W'(1 - \pi_J)} \\ &= \frac{h_\pi}{h_\pi + h_W}. \end{aligned} \tag{8}$$

Differentiating the CAF w.r.t. $\ln t$, we find that the CAF in de DHM is decreasing iff

$$(\ln h_\pi)' < (\ln h_W)'. \tag{9}$$

That is iff

$$\frac{\pi''_J}{\pi'_J} + \frac{\pi'_J}{1-\pi_J} < (\ln h_W)'. \tag{10}$$

If we, like Van Breukelen (1989), select for W the Weibull distribution, expression (10) becomes elegantly simple. So let the distribution $W = W(t)$ of T_W be the Weibull, and $w = w(t)$ its derivative w.r.t. $\ln t$,

$$\begin{aligned} W(t) &= 1 - \exp\left[-\left(\frac{t}{\beta}\right)^\gamma\right] \\ w(t) &= \frac{\gamma t^\gamma}{\beta^\gamma}\exp\left[-\left(\frac{t}{\beta}\right)^\gamma\right] = h_W(t)\,(1 - W(t))\,, \end{aligned} \tag{11}$$

with hazard function h_W given by

$$\ln h_W(t) = \ln\gamma - \gamma\ln\beta + \gamma\ln t, \tag{12}$$

where $\beta > 0$ is a scale parameter and $\gamma > 0$ a shape parameter. The larger γ the more probability mass at the lower values of t, so γ may be called a *time pressure parameter*. It follows that $(\ln h_W)' = \gamma$. Equation (10) implies that the CAF is decreasing iff

$$\frac{\pi''_J}{\pi'_J} + \frac{\pi'_J}{1-\pi_J} = \frac{(1-\pi_J)''}{(1-\pi_J)'} - \frac{(1-\pi_J)'}{1-\pi_J} < \gamma \tag{13}$$

It is easily checked that under assumptions A-1 and A-2 $\lim_{\vartheta\to\infty}$ of all three functions $1-\pi_J$, $(1-\pi_J)'$, and $(1-\pi_J)''$ equals zero, and application of l'Hôpital's rule shows that

$$\lim_{\vartheta\to\infty}\left(\frac{(1-\pi_J)''}{(1-\pi_J)'} - \frac{(1-\pi_J)'}{1-\pi_J}\right) = 0. \tag{14}$$

Equation (14) implies that there exists a smallest ϑ_d such that

$$\forall_{\vartheta > \vartheta_d} \left(\frac{(1-\pi_J)''}{(1-\pi_J)'} - \frac{(1-\pi_J)'}{1-\pi_J} < \gamma \right). \tag{15}$$

Therefore, if there exists a t_d with $\vartheta(t_d) > \vartheta_d$, then the CAF decreases for all $t > t_d$, because $\vartheta(t)$ is increasing in t. However, if one chooses a model for the precision function $\vartheta(t)$ that not necessarily increases without bound for $t \to \infty$, it is not certain that t_d exists, and the CAF may never be decreasing.

3 A Parametric Model for $\underline{\pi}$, and $\vartheta(t)$

In this section we discuss a specific functional form for $\underline{\pi}$, and $\vartheta(t)$. Let items $i = 1, ..., I$ be presented to subject v, $v = 1, ..., V$. After the items are administered we have the score vector $\underline{x}_v = (x_{v1}, ..., x_{vI})^T$, with $0 \leq x_i \leq J_i$, $J_i > 0$, and the response time vector $\underline{t}_v = (t_{v1}, ..., t_{vI})^T$. The precision function is modeled by

$$\vartheta_v(t) = \vartheta_v = \ln \xi_v t = \ln \xi_v + \ln t = \kappa_v + \ln t \tag{16}$$

where $\xi_v > 0$, is a latent variable that represents the mental power of subject v. Also, of course, $t > 0$. Thus we have chosen a precision function that is strictly increasing in t and increases without bound. Note that if $\exp(\vartheta)$ is taken as the precision function, ξ precisely gives the increase in precision per unit time. That is

$$\exp(\vartheta(t+1)) - \exp(\vartheta(t)) = (t+1)\,\xi - t\xi = \xi, \tag{17}$$

conform the definition in the Introduction. Let $\alpha_i > 0$, and $\underline{\eta}_i = (\eta_{i0}, ..., \eta_{iJ_i})$ be the parameters of item i. Then the probability π_{vij} for v to score in category j ($j \in \{0, ..., J_i\}$) of item i at time t is modeled by

$$\pi_{vij} = \pi_{vij}(t) = \pi_i(X_{vi} = j|t) \propto \exp(j\alpha_i\vartheta_v(t) + \eta_{ij}). \tag{18}$$

The parameter α_i determines the slope of the probability functions, and η_{ij} is a category parameter. To obtain an identified model set $\eta_{i0} = 0$, for all i, $\mathcal{E}(\kappa) = 0$, and $\mathcal{V}ar(\kappa) = 1$, where $\mathcal{E}(.)$ denotes expectation, and $\mathcal{V}ar(.)$ variance. If ϑ_v were a constant this model would be the Generalized Partial Credit Model (GPCM), (Muraki, 1992). Because here ϑ_v is a function of t, it can be viewed as a generalization of the GPCM.

The time pressure hazard function was given earlier (Formula 12). Given the specifications in Formulas (16), and (18), the hazard function $h_\pi(t)$ is given by:

$$h_{\pi vi}(t) = \frac{\pi'_{viJ_i}}{1-\pi_{viJ_i}}. \tag{19}$$

where

$$\pi'_{viJ_i} = \frac{\partial}{\partial \ln t}\pi_{viJ_i} = \alpha_i \pi_{viJ_i}\left(J_i - \tau_{vi}\right), \tag{20}$$

with

$$\tau_{vi} = \tau_{vi}(t) = \sum_j j\pi_{vij}(t), \tag{21}$$

the expected score of v for item i at time t. For a binary item $\tau_{vi}(t) = \pi_{vi1}(t) = \pi_{vi}(t)$, and Formula (19) takes the simple form

$$h_{\pi vi}(t) = \alpha_i \pi_{vi}(t). \tag{22}$$

To obtain a simple likelihood function, and to obtain enough data to estimate with acceptable accuracy the parameters of the time pressure hazard, h_W is assumed to depend only on the item index i. This implies that the time pressure is perceived equally among all subjects.

Applying Formula (13), a decreasing CAF results iff

$$a(t) = \frac{J_i - \tau_{vi}}{1-\pi_{viJ}} - \frac{\tau_{vi}^{[2]}}{J_i - \tau_{vi}} < \frac{\gamma_i}{\alpha_i}, \tag{23}$$

where (t_{vi}) is omitted to simplify the notation, and $\tau_{vi}^{[k]}(t)$ denotes k-th central moment of the item score at time t. If $t \to 0$, τ, π_{viJ}, and $\tau^{[2]}$ vanish, so that $\lim_{t\to 0} a(t) = J_i$. Although it is not in general true that $a(t)$ is decreasing in t, because its derivative may take positive values, it has been proven above that there exists a smallest t_d, such that $a(t) < \frac{\gamma_i}{\alpha_i}$ for $t > t_d$. This means that if $\lim_{t\to 0} a(t) = J_i > \frac{\gamma_i}{\alpha_i}$ then the CAF is initially increasing, and decreases for higher values of t. Indeed, some calculations show (see e.g. Figure 3) that for a broad range of parameter values $a(t)$ initially decreases smoothly towards

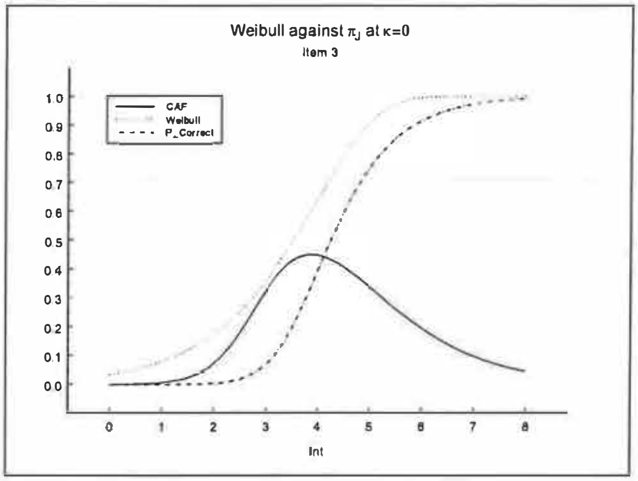

Figure 1: A Weibull, π_J and the resulting CAF at $\kappa = 0$

zero until at very high values of t, outside the range of relevant values, it may take small values around zero, remaining well below normal values of $\frac{\gamma_i}{\alpha_i}$. So it makes sense to calculate a value $t_{\max}$, for which $a(t_{\max}) = \frac{\gamma_i}{\alpha_i}$. Note that $t_{\max}$ depends on κ, because $a(t)$ does.

Examples of CAFs are shown in Figures 1 and 2. It is clear that the CAF at $\kappa = 2$ rises higher and starts earlier to decline than at $\kappa = 0$. In general, κ is a location parameter of π_J. Higher values of κ shift the graph to the left, thereby increasing the probability that $t_\pi < t_W$, and thus that a correct response and a shorter time results. With higher time pressure, which corresponds to increasing γ, the CAF also declines earlier. Figure 3 gives an impression of the relation between $a(t)$, its derivative and the CAF for an arbitrary item with $J = 3$. As can be seen $a(t)$ decreases smoothly towards zero.

Note that Formula (23) does not depend on β_i, and depends on t only through ϑ. The latter observation implies that one may solve numerically the equation implied by Formula (23) for ϑ, the solution being say ϑ^*. Then, using Formula (16) we have that

$$\vartheta^* = \kappa + \ln t_{\max}(\kappa), \tag{24}$$

and so

$$\exp(\vartheta^*) = t_{\max}(0) \tag{25}$$

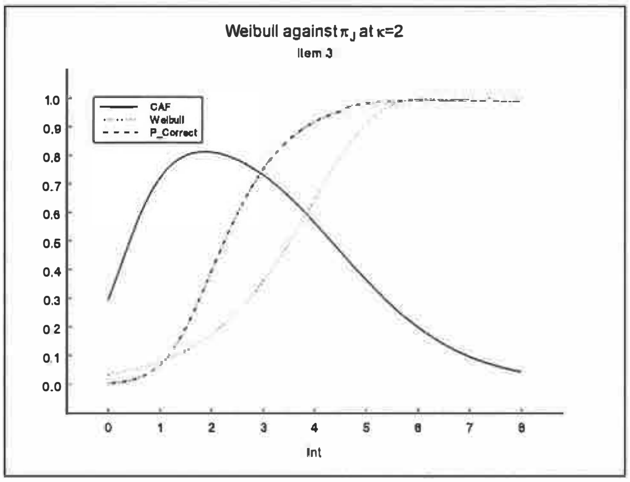

Figure 2: A Weibull, π_J and the resulting CAF at $\kappa = 2$

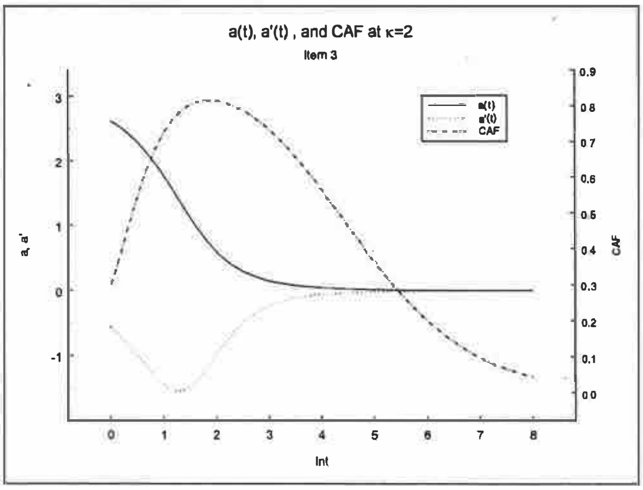

Figure 3: a(t), a'(t) (left axis), and CAF(t) (right axis) at $\kappa = 2$

For other values of κ, we have

$$t_{\max}(\kappa) = \exp(\vartheta^*)\exp(-\kappa) = t_{\max}(0)\exp(-\kappa). \tag{26}$$

For a binary item $\tau^{[2]}_{vi} = \pi_{vi}(1-\pi_{vi})$, and Formula (23) takes the simple form

$$1 - \pi_{vi}(t) < \frac{\gamma_i}{\alpha_i}, \tag{27}$$

which means that for binary items the CAF is always decreasing if $\gamma_i > \alpha$. Otherwise, the CAF is increasing up to a time

$$t_{\max vi} = \pi^{-1}_{vi}\left(\frac{\alpha_i - \gamma_i}{\alpha_i}\right) = \exp\left(\frac{\ln\left(\frac{\alpha_i-\gamma_i}{\gamma_i}\right) - \alpha_i\kappa - \eta_i}{\alpha_i}\right), \tag{28}$$

and decreasing afterwards.

All these properties are according to the model, and with unlimited observations. However, even if $t_{\max vi} > 0$, in practice there will be no observations at $t = 0$ and immediately afterwards, the first observations occurring from t_{is}, say. It may happen that $t_{is} > t_{\max vi}$ for all subjects. In that case only the decreasing part of their CAF's is shown by the data. Conversely, if the last observations occur before $t_{\max vi}$, the data only show the increasing part of the CAFs. If the observations occur around $t_{\max vi}$ and the CAFs are relatively flat, the data may suggest time independent CAFs. Moreover, as already mentioned, $t_{\max vi}$ depends on κ_v, which means that observations have to be from a group with homogeneous κ to possibly show a clear form of a CAF. Otherwise increasing and decreasing parts of CAFs tend to be pooled, which prevents a clear picture to emerge.

In the following section an MML estimation procedure of the model parameters is discussed.

4 Estimation

4.1 Iterative algorithms

In this section an EM-algorithm (Dempster, a.o., 1977) for the MML-estimation of the logistic model parameters is developed. It will appear that the parameters of the time pressure distribution, whatever its choice, Weibull, lognormal, etc., can be separately estimated with the Newton Raphson algorithm. We start with the simultaneous likelihood for all model parameters.

Consider κ_v as missing data with density g. The data generating process as introduced in Section 2 together with g determine the complete loglikelihood of the response matrix $(\underline{x}_v, \underline{t}_v)$ and κ_v. Let $I_{=J_i} \doteq I(x_{vi} = J_i)$ be the indicator function that takes the value 1 if $x_{vi} = J_i$, and 0 otherwise. Likewise $I_{<J_i} \doteq I(x_{vi} < J_i)$. Take $\kappa_v = \ln \xi_v$ as the subject parameter, and denote the vector of model parameters as $\underline{\lambda} = (\underline{\lambda}_\pi, \underline{\lambda}_W)$, with $\underline{\lambda}_\pi$ the logistic item parameters, and $\underline{\lambda}_W$ the Weibull item parameters. The complete data loglikelihood is then given by

$$\begin{aligned}
\ell_v(\underline{\lambda}; \underline{x}_v, \underline{t}_v, \kappa_v) &= \sum_i I_{<J_i} \ln \left(w_i \left(1 - \pi_{viJ_i}\right) \frac{\pi_{vix_{vi}}}{1 - \pi_{viJ_i}} \right) + \\
&\quad I_{=J_i} \ln \left(\pi'_{viJ_i} \left(1 - W_i\right)\right) + \ln g(\kappa_v) \\
&= \sum_i I_{<J_i} \ln \left(w_i \pi_{vix_{vi}}\right) + I_{=J_i} \ln \left(\left(1 - W_i\right) \pi'_{viJ_i}\right) + \ln g(\kappa_v)
\end{aligned} \tag{29}$$

Using the last part of (11), and (20), and rearranging terms it is found that

$$\begin{aligned}
\ell_v(\underline{\lambda}; \underline{x}_v, \underline{t}_v, \kappa_v) &= \sum_i \ln \pi_{vix_{vi}} + I_{=J_i} \ln \left(J_i - \tau_{vi}\right) + I_{=J_i} \ln \alpha_i + \\
&\quad \ln \left(1 - W_i\right) + I_{<J_i} \ln h_W + \ln g(\kappa_v) \\
&= \ln y_1(\underline{\alpha}, \underline{\eta}; \underline{x}_v, \underline{t}_v | \kappa_v) + \ln y_2(\underline{\alpha}; \underline{x}_v) + \\
&\quad \ln y_3(\underline{\beta}, \underline{\gamma}; \underline{x}_v, \underline{t}_v) + \ln g(\kappa_v) + D
\end{aligned} \tag{30}$$

with

$$\begin{aligned}
y_{1v}(\underline{\alpha}, \underline{\eta}; \underline{x}_v, \underline{t}_v | \kappa_v) &= \prod_i \pi_{vix_{vi}} \left(J_i - \tau_{vi}\right)^{I_{=J_i}} \\
y_{2v}(\underline{\alpha}; \underline{x}_v) &= \prod_i \alpha_i^{I_{=J_i}} \\
y_{3v}(\underline{\beta}, \underline{\gamma}; \underline{x}_v, \underline{t}_v) &= \prod_i \left(1 - W_i\right) \left(h_{Wi}\right)^{I_{<J_i}}
\end{aligned} \tag{31}$$

and D a constant depending only on the data. The complete likelihood, therefore, factors into four factors, of which only the first and the last depend on κ. Notice in particular that the likelihood of the Weibull parameters of each item appear as a separate factor of the likelihood, while we did not use a property of the Weibull itself. This entails two conclusions. First the Weibull parameters can be estimated per item separately from the logistic parameters. Secondly, this result is independent of the Weibull, and would have been obtained with any other choice for the distribution for time pressure.

The loglikelihood ℓ_{Wi} of the Weibull parameters for item i is given by

$$\ell_{Wi} = \sum_v \ln y_{3v} = \sum_v I_{<J_i} \ln h_W - \sum_v \left(\frac{t}{\beta}\right)^\gamma, \tag{32}$$

which can be maximized by the Newton-Raphson algorithm.

Concentrating only on the logistic item parameters $\underline{\lambda}_\pi$, one may omit y_3 from the likelihood. The marginal loglikelihood $M\ell_\pi$ of the logistic item parameters can then be written as

$$M\ell_\pi + C = \sum_v \ln y_{2v} + \int \ln y_{1v}(\underline{\lambda}_\pi; \underline{x}_v, \underline{t}_v|\kappa) h(\kappa|\underline{x}_v, \underline{t}_v) d\kappa, \tag{33}$$

where C is some constant, and h, the posterior distribution of κ, is given by

$$\begin{aligned} h(\kappa; \underline{\lambda}_\pi|\underline{x}_v, \underline{t}_v) &= \frac{y_1(\underline{x}_v; \underline{\lambda}_\pi|\kappa, \underline{t}_v) g(\kappa)}{\int y_1(\underline{x}_v; \underline{\lambda}_\pi|\kappa, \underline{t}_v) g(\kappa) d\kappa} \\ &= \frac{\prod_i \pi_{vix_{vi}} (J_i - \tau_{vi})^{I=J_i} g(\kappa)}{\int \prod_i \pi_{vix_{vi}} (J_i - \tau_{vi})^{I=J_i} g(\kappa) d\kappa}. \end{aligned} \tag{34}$$

The function that is to be maximized iteratively in the EM-method, derives from the posterior expected loglikelihood, and is given by

$$\begin{aligned} Q(\underline{\lambda}_\pi, \underline{\lambda}_\pi^*) &= \sum_v \ln y_{2v}(\underline{\alpha}) + \\ & \int \ln y_{1v}(\underline{\lambda}_\pi|\kappa) h_v(\kappa; \underline{\lambda}_\pi^*) d\kappa, \end{aligned} \tag{35}$$

with $\underline{\lambda}_\pi^*$ the parameter values from the previous iteration. The Appendix provides further details.

4.2 Initial Values

Here attention will be given to the problem of finding initial values for the iterative estimation algorithms for the Weibull and the logistic distributions. To simplify the notation we omit the item index. Let

$$\overline{\ln t} = \frac{\sum_v \ln t_v}{V} \tag{36}$$

denote the mean over subjects of $\ln t$.

First initial values for the logistic parameters are derived. Assuming that $\kappa = 0$, it follows from Formulas (16), (??), (18), and (29) that for $j < J$

$$\eta_j \approx \ln \frac{N_j + 0.5}{N_0 + 0.5} - j\alpha\overline{\ln t} \tag{37}$$

where N_j denotes the number of subjects who responded in category j, and 0.5 is added to prevent division by zero. Similarly, for $j = J$ we have that

$$\begin{aligned} \eta_J \quad &\approx \quad \ln \frac{N_J + 0.5}{N_0 + 0.5} - J\alpha\overline{\ln t} + \\ &\quad \left[\gamma\left(\overline{\ln t} - \ln\beta\right) + \ln\gamma - \ln(J - \bar{x}(\exp(\overline{\ln t})) - \ln\alpha\right]. \end{aligned} \tag{38}$$

It appears from calculations that the second part of Formula (38) does not have a great impact. So for practical purposes it may be neglected. The value of α does have a great impact, however. Fortunately, given initial values for $\underline{\eta}$, the EM-estimation procedure appears robust enough to cope with a common initial value for α like 1.0 for all items.

Initial values for the Weibull parameters can be found as follows. Let

$$\begin{aligned} Lt_< &= \sum_v I_{<J_i} \ln t_v \\ N_< &= \sum_v I_{<J_i} \end{aligned} \tag{39}$$

then equate the first derivatives of Formula (32) w.r.t. (β, γ) to zero (see also the Appendix), and substitute for $\beta^{-\gamma}$ in the expression for γ. This yields

$$\begin{aligned} \beta^{-\gamma} &= \frac{N_<}{\sum_v t^\gamma} \\ \gamma^{-1} &= \frac{\sum_v t^\gamma \ln t}{\sum_v t^\gamma} - \frac{Lt_<}{N_<} \end{aligned} \tag{40}$$

Now

$$S = \frac{\sum_v t^\gamma \ln t}{\sum_v t^\gamma} \tag{41}$$

is a weighted mean of $\ln t$ with higher weights for higher $\ln t$. Therefore, $S \approx \overline{\ln t} + c \times sd_{\ln t}$, for some constant $c > 0$, and $sd_{\ln t}$ the standard deviation of $\ln t$. From some trials it appeared that $c = 1.0$ yields reasonably accurate estimates of (β, γ). Using this approximation for S and Formula (40) first an initial value for γ is calculated, and then, using the initial value for γ, the initial value for β is found.

5 Testing the Model

Model tests can be constructed using the framework of the *Lagrange Multiplier* (LM) test-statistic. An introduction to the LM-test within a larger context can be found in Buse (1982). The idea for the LM-test originates with Rao (1948), there called the 'score test', and with Aitchinson and Silvey (1958). An application within the context of IRT models can be found in Glas and Verhelst (1995), and Glas (1997, 1999). In general, to compute the LM-statistic restrictions on parameters are relaxed. For instance, one may for a certain item i release the restriction that α_i is equal for boys and girls, thereby replacing α_i with α_{ib} for the boys, and α_{ig} for the girls. The restriction states that $\alpha_{ib} = \alpha_{ig} = \alpha_i$. Let superscript T denote transposition, then the LM-test statistic can be expressed as

$$LM = \ell^{(1)T} \left(-\ell^{(2)}\right)^{-1} \ell^{(1)}, \tag{42}$$

where the superscripts within parentheses denote order of differentiation with respect to the parameter-vector in the relaxed model, evaluated at the maximum likelihood estimate of the restricted model. LM is χ^2-distributed with degrees of freedom equal to the number of relaxed restrictions. E.g., in the example there is one released restriction $\alpha_{ib} = \alpha_{ig}$, and if LM for this restriction is significant one may conclude that the data do not support it.

To obtain an especially simple procedure for the calculation of the LM statistic one keeps all original parameters in the extended model, and changes the status of implicit (0 or 1) or explicit constants from constant to a variable parameter in the likelihood function. In general this is not the case. In the example above α_i was replaced with α_{ib} , and α_{ig}, where the original α_i disappeared from the model. In the sketched approach, one replaces α_i for instance by $\alpha_i + \alpha_{ib}$, and $\alpha_i + \alpha_{ig}$.

Denote the original parameters by $\underline{\lambda}$, the U new parameters by $\underline{\psi}$, and the complete vector of original and new parameters by $\underline{\zeta} = (\underline{\lambda}, \underline{\psi})$. The likelihood function is then evaluated at the maximum likelihood estimate of $\underline{\lambda}$, and of $\underline{\psi}$ at its restricted value (e.g. $\alpha_{ib} = 0$, and $\alpha_{ig} = 0$). Because the likelihood is evaluated at the maximum likelihood estimates of $\underline{\lambda}$, the elements of the first derivative w.r.t. $\underline{\lambda}$ are all equal to zero. Because $\underline{\lambda}$ remains completely in the relaxed model this simplifies the computation. Select with $F(\underline{\lambda})$ the vector of elements of the first derivatives of $M\ell$ with respect to the elements of $\underline{\lambda}$. Likewise $I(\underline{\lambda}, \underline{\psi})$ selects the part of the observed information matrix $I(= -M\ell^{(2)})$ with the rows for $\underline{\lambda}$, and the columns for $\underline{\psi}$. Then

$$LM(\underline{\psi}) = F(\underline{\psi})^T W^{-1} F(\underline{\psi}), \tag{43}$$

with

$$W = I(\underline{\psi}, \underline{\psi}) - I(\underline{\psi}, \underline{\lambda}) I(\underline{\lambda}, \underline{\lambda})^{-1} I(\underline{\lambda}, \underline{\psi}) \tag{44}$$

where $I(\underline{\lambda}, \underline{\lambda})^{-1}$ is already computed to obtain standard errors of the parameter estimates. If original parameters are replaced by new parameters by relaxation of restrictions, this simplification is not obtained.

In case the LM-test shows that the model is violated for a certain restriction, one may evaluate the size of the misfit with the first Newton-Raphson (N-R) step of the new parameters, were the estimation continued after releasing the restrictions. This first step is given by $F(\underline{\psi})^T W^{-1}$.

As to the construction of the test-statistics we follow the framework for item-oriented test statistics developed by Glas (1999). Adapted to the present model the procedure runs as follows. Drop the restriction that for a certain item i some or all parameters $\underline{\lambda}_i$ are independent of mental power κ. Divide the subjects into groups g ($g = 1, ..., G$) of homogeneous values for κ, and introduce new parameters $\underline{\psi}_{ig}$, so that the parameters for item i in the unrestricted model are $\underline{\lambda}_i + \underline{\psi}_{ig}$, depending on group membership g. For instance the model given in Formula (18) can be reformulated as

$$\pi_{vijg} \propto \exp\left(j\left(\alpha_i + \alpha_{ig}\right)\vartheta_{vi} + \eta_{ij} + \eta_{ijg}\right) \tag{45}$$

given that v is a member of mental power group g. In the original model it was assumed that $\alpha_{ig} = \eta_{ijg} = 0$.

Using a marginal model the EAP is commonly used as an estimate of the person parameter, and therefore indicates group membership. However, to obtain a simple formulation of the likelihood for the unrestricted model the estimate must be independent of the response to item i. Fortunately, the EAP can be cheaply obtained for each $\underline{x}^{(i)}$ separately, where $\underline{x}^{(i)}$ denotes the response vector $\underline{x}$ without the response to item i. Using Formula (34) the posterior $h(\kappa_q|\underline{x})$ can be approximated as

$$h(\kappa_q|\underline{x}) \approx \frac{w_q \prod_i f(x_i;\kappa_q)}{\sum_q w_q \prod_i f(x_i;\kappa_q)} \tag{46}$$

where w_q $(q = 1, ..., Q)$ is the Gauss-Hermite weight at κ_q. Now consider the $Q-$vector $\underline{y}$ with values

$$y_q = w_q \prod_i f(x_i;\kappa_q) \tag{47}$$

then the posterior distribution of κ at κ_q given $\underline{x}^{(i)}$ is given by

$$h(\kappa_q|\underline{x}^{(i)}) \propto z_{iq} = \frac{y_q}{f(x_i;\kappa_q)} \tag{48}$$

and the $EAP(\underline{x}^{(i)})$ is found with

$$EAP(\underline{x}^{(i)}) = \frac{\sum_q \kappa_q z_{iq}}{\sum_q z_{iq}}. \tag{49}$$

It is, of course, more simple when the group membership of respondents is given by a background variable like sex, or cultural environment.

It is a disadvantage of the proposed formulation of the relaxed model

$$\pi_{vijg} \propto \exp\left(j\,(\alpha_i + \alpha_{ig})\,\vartheta_{vi} + \eta_{ij} + \eta_{ijg}\right) \tag{50}$$

that an indeterminacy problem is introduced, because for some constant c,

$$\alpha_i + \alpha_{ig} = \alpha_i - c + \alpha_{ig} + c = \alpha_i^* + \alpha_{ig}^*. \tag{51}$$

Consequently, if the parameters $\alpha_i, \alpha_{i1}, ..., \alpha_{iG}$ are unrestricted the combination is undetermined and the information matrix cannot be inverted unless a linear restriction S is imposed on $(\alpha_i,\underline{\alpha_i})$, or, in order to let the original $\underline{\lambda}$ untouched, a linear restriction T on $\underline{\alpha_i}$ suffices. We will take

$$\sum_g \alpha_{ig} = \sum_g \eta_{ijg} = 0. \tag{52}$$

This choice has the advantage that a new parameter step is calculated for each group, without having to recalculate an inverse of a different information matrix of original parameters for each item.

Because α_i is a scale parameter, a multiplicative correction $\alpha'_{ig}\alpha_i$ might seem more appropriate, and have a clearer interpretation across items. However, the additive group parameters for α_i introduced by Glas (1999) have the advantage that the derivatives w.r.t. α_i and α_{ig} in the new model retain the same form for $\alpha_{ig} = 0$ as the derivative w.r.t. α_i in the original model. Both advantages can be obtained by introducing the reparameterization $\rho = \ln \alpha$, with

$$\pi_{vijg} \propto \exp\left(j\left(\exp(\rho_i + \rho_{ig})\,\vartheta_{vi} + \eta_{ij} + \eta_{ijg}\right)\right), \tag{53}$$

and taking derivatives w.r.t. ρ. In the same vein we have for the Weibull parameters

$$\ln h_{Wvig} = \ln(\gamma_i + \gamma_{ig}) + (\gamma_i + \gamma_{ig}) \ln \frac{t_{vi}}{\beta_i + \beta_{ig}} - \ln t_{vi}, \tag{54}$$

with

$$\sum_g \gamma_{ig} = \sum_g \beta_{ig} = 0 \tag{55}$$

The conditional probability for v to obtain response vector $(\underline{x}_v, \underline{t}_v)$ on item i in the relaxed DHM given that $EAP(\underline{x}_v^{(i)}, \underline{t}_v^{(i)})$ classifies v into group g, denoted as $EAP(\underline{x}_v^{(i)}, \underline{t}_v^{(i)}) \in g$, is then given by

$$P(\underline{x}_v, \underline{t}_v; \kappa_v, \underline{\zeta} | EAP(\underline{x}_v^{(i)}, \underline{t}_v^{(i)}) \in g) =$$
$$P(\underline{x}_v^{(i)}, \underline{t}_v^{(i)}; \kappa_v, \underline{\lambda}^{(i)}) P(x_{vi}, t_{vi}; \kappa_v, \underline{\zeta}_i | EAP(\underline{x}_v^{(i)}, \underline{t}_v^{(i)}) \in g).$$

Glas (1999) following Mislevy (1986) mentions that, for the purposes of LM-tests, the observed information matrix can be replaced by an approximation of the Fisher information matrix.

$$A=\sum_{v}\mathcal{E}_{\kappa}\left(\frac{\partial}{\partial\underline{\pi}}\ell\left(\underline{\pi};\underline{x}_{v}\right)\right)\mathcal{E}_{\kappa}\left(\frac{\partial}{\partial\underline{\pi}'}\ell\left(\underline{\pi};\underline{x}_{v}\right)\right) \tag{56}$$

where $\underline{\pi}=(\underline{\xi},\underline{\lambda})$. Here follows another proof of this approximation.

Proof: denote $\frac{\partial}{\partial\zeta_i}$ as ∂_i, and let $\mathcal{E}_{\underline{x}}f(\underline{x})$ denote the expectation of $f(\underline{x})$ over the distribution of $\underline{x}$. For a general loglikelihood function $\ell^*\left(\underline{\zeta};\underline{x}\right)$ it holds that

$$\begin{aligned}-\mathcal{E}_{\underline{x}}\partial_i\partial_j\ell^* &= \mathcal{E}_{\underline{x}}\partial_i\ell^*\partial_j\ell^* \\ &\approx \frac{1}{V}\sum_{v}\partial_i\ell_v^*\partial_j\ell_v^* \end{aligned} \tag{57}$$

For a marginal loglikelihood $M\ell\left(\underline{\zeta};\underline{x}_v\right)=\mathcal{E}_{\kappa}\left(\ell\left(\underline{\zeta};\underline{x}_v\right)\right)$we have that $\partial_i M\ell=\mathcal{E}_{\kappa}\left(\partial_i\ell\right)$. Substitution of $M\ell$ for ℓ^* in Formula (57) gives Formula (56).■

In the present context, the greater generality of this proof allows to apply it not only to the marginal loglikelihood for the logistic parameters, but also to the ordinary loglikelihood for the Weibull parameters.

Above, it was mentioned that $I(\underline{\lambda},\underline{\lambda})^{-1}$ was already computed to obtain standard errors for the parameter estimates. It is of course tempting to use this result in the computation of the Langrange Multiplier tests. However, from some worked examples it appeared that, using the approximation (56) for $I(\underline{\psi},\underline{\lambda})$, and $I(\underline{\psi},\underline{\psi})$ in combination with $I(\underline{\lambda},\underline{\lambda})$, computed with the method of Louis (1982), it may happen that noninvertible matrices W result. This problem does not occur if the entire information matrix, including the part $I(\underline{\lambda},\underline{\lambda})$, is calculated using approximation (56).

6 An Illustration

An example of the results on a simulated data set analyzed with the DHM is shown in the tables below. The data were generated according to the DHM with 20 items with $J=3$ in a complete design with 300 records. The distribution of κ was the standard normal. All items had the same Weibull parameters $(\beta,\gamma=50,0.80)$. The discrimination parameters were 0.3, 1.0, 1.5, distrubuted over the items in about equal amounts. The category parameters were chosen so as to avoid low category frequencies. It turned out that $\eta_{ij}=-\frac{j}{2}\alpha_i\ln\mathcal{E}T_{Wi}$, with $\mathcal{E}T_{Wi}=\beta_i\Gamma((\gamma_i+1)/\gamma_i)$, the expected response time from the Weibull distribution, was a proper choice. Two items were excepted from this scheme. Via items 1 and 3 a model violation was introduced for which the LM-tests should be sensitive. We had $\alpha_1 = 1.0$ for $\kappa < 0$ and $\alpha_1 = 2.0$ for $\kappa \geq 0$. For item 3 we had $\gamma_3 = 0.6$ for $\kappa < 0$, and $\gamma_3 = 1.2$ for $\kappa \geq 0$.

Table 1 shows some general statistics, like the mean response time of an item and its standard deviation, and the number of observations on an item and its categories. Table 2 displays some results of the parameter estimation. The number of mentioned iterations (60) only refers to the EM-algorithm for the logistic parameters. The Weibull parameters are estimated with a few Newton-Raphson iterations. The original values of the parameters used to generate the data are shown in the column headed Orig.Val. The column headed FirstD contains the first derivatives of Q at the last iteration. TMax(0) gives the response time where the CAF has its modal value for $\kappa = 0$, and MnX gives the mean item score for $\kappa = 0$ at TMax(0). The first LM-statistic (LMStat) in the row labeled 'α' in Table 3, shows the LM-test statistic where only the constraint on the discrimination index is relaxed. The second LMStat is calculated relaxing all constraints of the logistic group parameters $(\alpha_g, \underline{\eta}_g)$ of the item. The NR-step for the discrimination parameter refers to α, not to $\rho = \ln \alpha$. Therefore, with two groups we have that $\psi_1 = -\psi_2$ for all parameters, and the sum over groups of the NR-step equals zero, except for α, because the sum for ρ equals zero. Finally, Table 4 displays the LM-statistics for the Weibull parameters in the same way as Table 3 for the logistic parameters.

It appears from Table 2 that the parameter estimates are close to their original values as measured by their Standard Errors. Except, of course, for the model violations. Unexpectedly the estimate of α_1 is even less than its original minimum value. This also has its impact on the estimates for $\underline{\eta}_1$ in relation to their original values. As the Weibull estimates are independent of the logistic parameters, β_1 and γ_1 are accurately estimated. The same insensitivity can be observed at item 3 but this time for the logistic parameter estimates, who are not affected by the model violation of γ_3. The underestimation also holds for γ_3, although less conspicuous. Its estimate barely raises above its lowest original value.

Item 2 has a declining CAF for $\kappa = 0$ from the start, or almost from the start, because its $t_{\max}(0) < 0.5$. One should realize that we did not choose a unit of time. Indeed, this is rather immaterial, except for the interpretation of $t_{\max}(0)$ with real data. As Formula (16) shows, a change of unit only adds a constant c to the precision parameter ϑ, which is compensated by adding $-\alpha jc$ to η_j.

Table 3 shows that the model violation of α_1 is very clearly detected by the LM-test. The LM-test for only the restriction on α_1 has a value of 129.9 with 1 degree of freedom, p= 0.00000. For the restrictions of all parameters of item 1 the value of the LM-test is slightly higher, indicating that the culprit must be the restriction on α_{1g}. The model violation of γ_3 is less clearly detected, but still shows a highly significant LM-test of 9.4 at two degrees of freedom, p= 0.009. This statistic is calculated simultaneously for the restrictions on both Weibull parameters. Therefore, the LM-statistic by itself cannot inform whether the restrictions on β_{3g} or γ_{3g} or both are to be blamed. However, it can be inferred from the first derivatives per group, which are relatively high for γ_3, that the restriction on γ_{3g} are the cause for a high LM-statistic for the Weibull parameters of item 3. The NR-step size for β_3 may seem appreciable (17.03). However, compared to its standard error (12.51), its relative size (1.36) is less than the relative size of the NR-step for γ_3 $(0.10/0.05 = 2.0)$.

The LM-statistic in its present form necessitates to divide the respondents into homogeneous mental power groups. When respondents incline to spend equal amounts of time on average on items, this division also produces homogeneous precision within groups. This results in unequal frequencies per category within groups, for instance in the low mental power group a higher frequency for low categories than for high categories. Especially if one decides to divide in more than two groups, the lower frequencies might affect the desired asymptotic behavior of the test statistics.

Table 1

Mean time and observations per item score

ItId	Mean time	Sd time	J	N	N per score			
1	24.6	7.6	3	300	69	24	37	170
2	33.4	8.0	3	300	38	62	76	124
3	18.1	6.8	3	300	45	17	40	198
4	26.3	7.5	3	300	37	51	80	132
5	26.9	7.5	3	300	45	46	70	139
6	18.2	6.5	3	300	45	36	44	175
7	30.3	7.6	3	300	35	55	81	129
8	25.7	7.4	3	300	33	44	69	154

Table 2
Calibration results after 60 EM-iterations

ItId	Par	N	Orig.Val	Estim.	St.Err.	FirstD	TMax(0)/MnX
1	β		50.0	57.56	6.43	0.000	9.26
	γ		0.80	0.79	0.05	-0.000	1.62
	α	69	1.0, 2.0	0.75	0.06	-0.009	
	η_1	24	-3.03	-2.42	0.25	0.000	
	η_2	37	-6.05	-3.88	0.27	0.000	
	η_3	170	-9.08	-4.82	0.38	-0.002	
2	β		50.0	55.04	5.00	0.000	<0.50
	γ		0.80	0.83	0.05	-0.000	0.88
	α	38	0.3	0.32	0.02	-0.001	
	η_1	62	-0.61	-0.18	0.21	0.001	
	η_2	76	-1.21	-0.99	0.21	-0.000	
	η_3	124	-1.82	-1.77	0.24	-0.001	
3	β		50.0	72.13	12.51	0.000	17.51
	γ		0.6, 1.2	0.66	0.05	0.000	2.49
	α	45	1.0	1.22	0.10	-0.012	
	η_1	17	-2.02	-2.10	0.35	0.001	
	η_2	40	-4.04	-4.11	0.37	0.001	
	η_3	198	-6.06	-6.98	0.60	-0.003	
...	...	...		...	...	...	...

Table 3
LM-statistics for 2 homogeneous κ-groups
Logistic parameters

ItId	Grp	Par	N	FirstD	NR-step	LM-stat	DF	P(χ^2>L)
1	1	α	58	-178.41	-0.33	129.93	1	0.00000
		η_1	20	9.89	-0.11	133.25	4	0.00000
		η_2	26	11.40	0.29			
	2	η_3	49	-39.83	1.02			
		α	11	178.40	0.60			
		η_1	4	-9.89	0.11			
		η_2	11	-11.40	-0.29			
		η_3	121	39.83	-1.02			
2	1	α	25	-2.41	-0.03	0.088	1	0.76029
		η_1	35	-1.86	0.01	3.025	4	0.55635
		η_2	38	0.25	0.17			
...	...	...	...	...	...	...	...	...

Table 4
LM-statistics for 2 homogeneous κ-groups
Weibull parameters

ItId	Grp	Par	N	FirstD	NR-Step	LMStat	DF	P(χ^2>L)
2	1	β	151	5.21	4.24	1.16	2	0.56694
		γ		-5.73	-0.03			
	2	β	149	-5.21	-4.24			
		γ		5.73	0.03			
3	1	β	158	-6.77	-17.03	9.44	2	0.00898
		γ		-21.54	-0.10			
	2	β	142	6.77	17.03			
		γ		21.54	0.10			
...	...	...	...	...	...	...	...	...

7 Discussion

From Formulas (18) and (16) it appears that the value of $\underline{\eta}$ depends on the response times. The longer the response times for item i, the lower (the more negative) the estimates of $\underline{\eta}_i$. This may seem an undesirable property for an item parameter. However, a reparameterization of the model may better emphasize the item parameters as a property of the item, just as independent of the response time as the subject parameter for mental power. Let ϕ_{ij} be defined by

$$\eta_{ij} = -j\alpha_i \ln(\phi_{ij}) \tag{58}$$

then

$$\begin{aligned} f_i(X_{vi} = j) &\propto \exp(j\alpha_i\vartheta_v + \eta_{ij}) \\ &= \exp(j\alpha_i(\ln\xi_v + \ln t_{vi} - \ln\phi_{ij})) \\ &= \left(\frac{\xi_v}{\phi_{ij}} t_{vi}\right)^{j\alpha_i} \end{aligned} \tag{59}$$

shows that $\underline{\phi}_i$ can be viewed as a scaling vector, that determines the impact of ξ_v with the advancement of time on the propensity-change of the categories of i. It, therefore, does not seem less reasonable to assume that this scaling property is sample invariant than to assume that the item parameters in precision models like the Rasch model are sample invariant. This reparameterization also supports the choice for the name 'power' parameter for $\xi = \exp(\kappa)$, because, as a physical analogue, it can be seen as the amount of energy per unit of time that has to be spent against the item-resistance vector $\underline{\phi}$ to effect a certain propensity change. The concept of 'speed' is a result of this confrontation, but is not itself represented in Formula (59).

It is a critical assumption in the DHM that if a person responds correctly, his response time is drawn from his π_J, and is shorter than his draw from the Weibull of the item. On the other hand, if he does not respond correctly, his π_J is assumed to be censored by the Weibull. This assumption creates much simplicity because it is known whether the Weibull was censored by π_J or not, and vice versa. It depends on this assumption that the parameters of the Weibull, or any other time distribution, appear in a separate factor of the likelihood. Moreover, on first sight it might appear that one could also attribute a positive probability $p(t)$ of a correct response when the response time is drawn from the Weibull. However, not only the known censoring is lost, but one also has to introduce a new function $g_{vi}(t) < \pi_{viJ}(t)$ $(g_{vi}(t) + p(t) = \pi_{viJ}(t))$ that describes the probability to enter the correct state. Consequently, there are compelling reasons to adhere to this assumption.

Finally, I want to spend some remarks on the relative value of mental power measurement. When all students follow more or less the same route towards the solution of a problem, power is undoubtedly a valuable ability worthy of professional evaluation. However, this condition is not always fulfilled. For instance, in solving high school physics problems, some students are very quick in correctly applying hardly understood formulas to superficially understood problems. Unfortunately, too often a successful strategy in the realm of school physics. Others, with a more thorough and conceptual understanding of physics may take more time to understand and solve these same problems. In a case like this, sheer uncritical power measurement would disadvantage the latter students, and, perhaps, unjustifiably convince some not to pursue a career in physics because of lack of talent. Therefore, I want to emphasize that results of power measurement should always be interpreted with due care and criticism.

8 Appendix: EM-Estimation in the Double Hazard Model

The normal distribution family is chosen as the population distribution $g(.)$. Integration is numerically approximated by Gauss-Hermite quadrature

$$\int_{\mathbb{R}} e^{-x^2} f(x)dx \approx \sum_{q=1}^{Q} w_q f(x_q) \tag{60}$$

where x_q, and w_q are calculated with a routine published in Press a.o. (1992, C, p. 154).

Let κ_q ($q = 1, ..., Q$) be the Gauss-Hermite points associated with $g(.)$ of the previous iteration, and

$h_{vq} = h(\kappa_q; \underline{\pi}^* \vert \underline{x}_v, \underline{t}_v)$	the Gauss-Hermite weight of the posterior density at κ_q given $\underline{x}_v$ and $\underline{t}_v$ evaluated at $\underline{\pi}^*$
$\pi_{vijq} = \pi_i(X_{vi} = j \vert \kappa_q, t_{vi})$	the probability of response j to item i at κ_q given response time t_{vi}, evaluated at $\underline{\pi}$.

Further, we define

$\tau_{viq} = \sum_k k\pi_{vikq}$	the expected score of item i at κ_q given response time t_{vi}
$d_{ij} = d_{vijq} = j - \tau_{viq}$	the deviation from the mean
$\tau^{[2]}_{viq}(x) = \sum_k \pi_{vikq} d^2_{vikq}$	the variance of the score of item i at κ_q given response time t_{vi}
$\varrho = \ln \alpha$	
$p_{i\varrho} = p = N(\mu^*_\varrho, \sigma^2_\varrho)$	the prior distribution of $\varrho = \ln \alpha$, μ^*_ϱ the mean of ϱ from the previous iteration, σ^2_ϱ is provided by the user
δ^{ij}	Kroneckers $\delta = 1$ if $i = j$, 0 otherwise
$I_{=J_i} = I(x_{vi} = J_i)$	The index function for correct responses
$I_{<J_i} = I(x_{vi} < J_i)$	The index function for incorrect responses
$N_< = \sum_v I_{<J_i} 1$	The number of incorrect responses (on i)
$Lt_< = \sum_v I_{<J_i} \ln t$	Sum of $\ln t$ on incorrect responses

To structure the derivations, the following formulas are useful, where the item index i is omitted. First on the logistic part.

$$\begin{aligned}
\pi'_{vxq} &= \frac{\partial}{\partial t}\pi_{vx}(t_q) = \frac{\alpha}{t_q}\pi_{vxq}d_{vxq} \\
\frac{\partial}{\partial \varrho}\ln \pi_{vxq} &= \alpha\vartheta_{vq}d_{vxq} \\
\frac{\partial}{\partial \varrho}\tau_{vq} &= \alpha\vartheta\tau^{[2]}_{vq} \\
\frac{\partial}{\partial \varrho}\tau^{[2]}_{vq} &= \alpha\vartheta_{vq}\sum_j \pi_{jvq}d^3_{vjq} = \alpha\vartheta\tau^{[3]}_{vq} \\
\frac{\partial}{\partial \eta_k}\ln \pi_{vxq} &= \delta_{kx} - \pi f_{vkq} \\
\frac{\partial}{\partial \eta_k}\tau_{vq} &= \pi_{vkq}d_{vkq} \\
\frac{\partial}{\partial \eta_k}\tau^{[2]}_{vq} &= \pi_{vkq}\left(d^2_{vkq} - \tau^{[2]}_{vq}\right)
\end{aligned} \tag{61}$$

next on the Weibull part

$$\begin{aligned}
h_W(t) &= \frac{\gamma}{\beta^\gamma}t^\gamma = \gamma\left(\frac{t}{\beta}\right)^\gamma \\
\ln h_W(t) &= \ln\gamma - \gamma\ln\beta + \gamma\ln t \\
&= \ln\gamma + \gamma\ln\frac{t}{\beta} \\
y_3 &= -\sum_v\left(\frac{t}{\beta}\right)^\gamma + \sum_v I_< \ln h_W \\
&= -\beta^{-\gamma}\sum t^\gamma + (\ln\gamma - \gamma\ln\beta)\,N_< + \gamma Lt_< \\
\frac{\partial}{\partial \ln\beta} - \beta^{-\gamma} &= \gamma\beta^{-\gamma} \\
\frac{\partial}{\partial \ln\gamma} - \beta^{-\gamma} &= \gamma\beta^{-\gamma}\ln\beta \\
\frac{\partial}{\partial \ln\gamma}t^\gamma &= \gamma t^\gamma \ln t
\end{aligned} \tag{62}$$

Especially the estimates of α are sensitive to chance capitalization. Therefore, a normal prior distribution $N(\mu_\varrho, \sigma^2_\varrho)$ for $\varrho = \ln\alpha$ is introduced, with μ_ϱ the mean of the current estimates of ϱ, and σ^2_ϱ the prior variance. The estimation of $\varrho = \ln\alpha$ instead of α, prevents divergence of the estimation algorithm as a result of oscillation between positive and negative values of α, or overshooting of the Newton-Raphson algorithm in the maximization step, and it enforces positive estimates. This last property may be viewed as a disadvantage. One has to be aware that an estimate of α close to zero may be indicative of an item with negative discrimination. The estimation accuracy for the other parameters will, in general, be high enough that a not too strict prior distribution will be overruled by the data.

Now we have that

$$\begin{aligned} Q(\underline{\lambda}, \underline{\lambda}^*) &\approx \sum_v \ln y_{2v}(\underline{\alpha}) + \sum_{v,q}\left(\sum_i \ln y_{1v}(\underline{\lambda}|\kappa_q) + \ln g(\kappa_q)\right) h_{vq} \qquad (63) \\ &= \sum_{vi} I_{=J_i}\varrho_i + \sum_{v,q}\left(\sum_i \ln \pi_{vix_{vi}} + I_{=J_i}\ln d_{viJq} + \ln g(\kappa_q)\right) h_{vq}. \end{aligned}$$

The first and second derivatives of $Q(.,.)$ are given below, where the item index, and sometimes other obvious indices are omitted. The part between $\{\}^*$ is added where appropriate and refers to the prior distribution on ϱ.

$$\begin{aligned} \frac{\partial}{\partial\varrho}Q &= \sum_v I_{=J_i}\frac{\partial}{\partial\varrho}\varrho + \sum_{vq} h_{vq}\left(\frac{\partial}{\partial\varrho}\ln\pi_{vxq} + I_{=J_i}\frac{\partial}{\partial\varrho}\ln d_J\right) \qquad (64) \\ &= N_J + \alpha\sum_{vq} h_{vq}\vartheta_{vq}\left(d_{x_v} - I_{=J_i}\frac{\tau^{[2]}_{vq}}{d_J}\right) = N_J + Q_\varrho \\ &= N_J + Q_\varrho\left\{-\frac{\varrho_i - \mu_\varrho}{\sigma^2_\varrho}\right\}^* \\ \frac{\partial}{\partial\eta_k}Q &= N_k - \sum_{vq} h_{vq}\pi_{vkq}\left(1 + I_{=J_i}\frac{d_k}{d_J}\right) = N_k - Q_k \\ \frac{\partial}{\partial\ln\beta}\ln y_3 &= \gamma\left(\beta^{-\gamma}\sum t^\gamma - N_<\right) \\ \frac{\partial}{\partial\ln\gamma}\ln y_3 &= N_< + \gamma\left[Lt_< - \beta^{-\gamma}\sum t^\gamma\ln t + \ln\beta\left(\beta^{-\gamma}\sum t^\gamma - N_<\right)\right] \\ &= N_< + \gamma\left[Lt_< - \beta^{-\gamma}\sum t^\gamma\ln t + \gamma^{-1}\ln\beta\frac{\partial}{\partial\ln\beta}\ln y_3\right] \end{aligned}$$

$$
\begin{aligned}
\left(\frac{\partial}{\partial\varrho}\right)^2 Q &= Q_\varrho - \alpha^2 \sum_{vq} h\vartheta^2 \left[\tau^{[2]} + I_{=J_i}\left(\frac{\tau^{[3]}}{d_J} + \left(\frac{\tau^{[2]}}{d_J}\right)^2\right)\right] \left\{-\frac{1}{\sigma_\varrho^2}\right\}^* \qquad (65)\\
\frac{\partial^2}{\partial\varrho\partial\eta_k} Q &= -\alpha \sum_{vq} h\vartheta\pi_k \left[d_k + \frac{I_{=J_i}}{d_J}\left(d_k^2 + (k-J)\frac{\tau^{[2]}}{d_J}\right)\right]\\
\frac{\partial^2}{\partial\eta_k\partial\eta_l} Q &= -\delta_{kl} Q_k + \sum_{vq} h\pi_k\pi_l \left(1 + \frac{I_{=J_i}}{d_J}\left(d_k - d_l\frac{k-J}{d_J}\right)\right)\\
\left(\frac{\partial}{\partial\ln\beta}\right)^2 \ln y_3 &= -\gamma^2\beta^{-\gamma}\sum_v t_v^\gamma\\
\frac{\partial^2}{\partial\ln\gamma\partial\ln\beta}\ln y_3 &= \frac{\partial}{\partial\ln\beta}\ln y_3 + \gamma^2\beta^{-\gamma}\left[\sum_v t_v^\gamma \ln t_v - \ln\beta\sum_v t_v^\gamma\right]\\
\left(\frac{\partial}{\partial\ln\gamma}\right)^2 \ln y_3 &= \frac{\partial}{\partial\ln\gamma}\ln y_3 - N_< + \gamma^2\beta^{-\gamma}\left[\begin{array}{c}\ln\beta\sum_v t_v^\gamma\ln t_v - \sum_v t_v^\gamma(\ln t)^2 + \\ \ln\beta\left(-\ln\beta\sum_v t_v^\gamma + \sum_v t_v^\gamma \ln t_v\right)\end{array}\right].
\end{aligned}
$$

$\frac{\partial^2}{\partial\eta_k\partial\eta_l}Q$ as given above does not seem symmetrical in k and l. However

$$
d_k - d_l\frac{k-J}{d_J} = d_k - d_l\frac{d_k - d_J}{d_J} = d_k + d_l - \frac{d_k d_l}{d_J}, \qquad (66)
$$

which is clearly symmetrical in k and l.

The estimation algorithm proceeds iteratively. Each iteration consists of an E-step and an M–step. In the E-step the derivatives of Q are calculated, and in the M-step the estimates of the parameters are improved by one Newton-Raphson step.

The asymptotic standard errors of estimation are calculated by finding the matrix of second derivatives of $M\ell$ with the method of Louis (1982).

9 References

Aitchinson, W. & Silvey, S.D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics, 28*, 813-828.

Breukelen, G.J.P. van (1989). *Concentration, speed and precision in mental tests; a psychonometric approach.* Thesis, Univ. of Nijmegen.

Breukelen, G.J.P. van (1995). Psychometric and information processing properties of selected response time models. *Psychometrika, 60*, 1, 95-113.

Breukelen, G.J.P. van (1997). Separability of item and person parameters in response time models. *Psychometrika, 62*, 4, 525-544.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. *The American Statistician, 36,* 3, part 1, 153-157.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Donders, A.R.Th. (1997). *The validity of basic assumptions underlying models for time-limit tests.* Design and analysis of speed-accuracy trade-off experiments. Thesis, KU Nijmegen: NICI.

Glas, C.A.W. (1997). Testing the generalized partial credit model. In M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.). *Objective measurement: theory into practice, Vol 4.* New Jersey: Ablex Publishing Corporation.

Glas, C.A.W. (1999). Modification indices for the 2-pl and the nominal response model. Accepted for publication in *Psychometrika.*

Glas, C.A.W. & Verhelst, N.D. (1995). Tests of fit for polytomous Rasch models. In G.H. Fischer & I.W. Molenaar (Eds.). *Rasch models, their foundation, recent developments and applications.* New York: Springer.

Louis, Th.A. (1982). Finding the information matrix when using the EM-algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika, 58*, 3, 445-469.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*, 177-195.

Muraki, E. (1992). A generalized partial credit model: application of an EM-algorithm. *Applied Psychological Measurement, 16*, 159-176.

Pachella, R.G. (1974). The interpretation of reaction time in information processing research. In B.H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Hillsdale, NJ: Erlbaum.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1992). *Numerical recipes in Pascal.* Cambridge: Cambridge University Press.

Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society, 44*, 50-57.

Roskam, E.E. (1987). Toward a psychometric theory of intelligence. In E.E.Ch.I. Roskam & R. Such (Eds.) *Progress in mathematical psychology.* Amsterdam: North Holland.

Roskam, E.E. (1997). Models for speed and time-limit tests. In W.J. van der Linden, and R.K. Hambleton (Eds.), *Handbook of modern item response theory.* New York: Springer.

Scheiblechner, H. (1979). Specifically objective stochastic latency mechanisms. *Journal of Mathematical Psychology, 19*, 18-38.

Scheiblechner, H. (1985). Psychometric models for speed-test construction: the linear exponential model. In S.E. Embretson (Ed.) *Test design, developments in psychology and psychometrics.* London: Academic Press.

Storms, L. & Delbeke, L. (1992). The irrelevance of distributional assumptions on reaction times in multidimensional scaling of same/different judgement tasks. *Psychometrika, 57*, 4, 599-614.

Verhelst, N.D., Verstralen, H.H.F.M. & Jansen, M.G.H. (1997). A logistic model for time-limit tests. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory.* New York: Springer.