



# Optimal designs for the generalized partial credit model

Paul-Christian Bürkner<sup>1\*</sup> , Rainer Schwabe<sup>2</sup> and Heinz Holling<sup>1</sup>

<sup>1</sup>Institute of Psychology, University of Münster, Germany

<sup>2</sup>Institute of Mathematical Stochastics, University of Magdeburg, Germany

Analysing ordinal data is becoming increasingly important in psychology, especially in the context of item response theory. The generalized partial credit model (GPCM) is probably the most widely used ordinal model and has found application in many large-scale educational assessment studies such as PISA. In the present paper, optimal test designs are investigated for estimating persons' abilities with the GPCM for calibrated tests when item parameters are known from previous studies. We find that local optimality may be achieved by assigning non-zero probability only to the first and last categories independently of a person's ability. That is, when using such a design, the GPCM reduces to the dichotomous two-parameter logistic (2PL) model. Since locally optimal designs require the true ability to be known, we consider alternative Bayesian design criteria using weight distributions over the ability parameter space. For symmetric weight distributions, we derive necessary conditions for the optimal one-point design of two response categories to be Bayes optimal. Furthermore, we discuss examples of common symmetric weight distributions and investigate under what circumstances the necessary conditions are also sufficient. Since the 2PL model is a special case of the GPCM, all of these results hold for the 2PL model as well.

## 1. Introduction

Item response theory (IRT) provides a flexible and powerful approach for designing and analyzing data of psychological tests. While IRT very often deals with dichotomous responses (e.g., categorized as either 'right' (1) or 'wrong' [0]), there are many situations where a more advanced scoring method is advisable. Consider, for instance, a complicated task in which multiple intermediate results have to be obtained to correctly solve the whole task. Scoring such a task in a dichotomous way comes with substantial loss of information as someone solving all intermediate questions and failing only in the last step would receive the same score as someone who did not even manage to take the first step. A natural solution is to give persons *partial credit* for the intermediate results. The response obtained is no longer dichotomous but ordinal, ranging from 0 (nothing correct), via 1 (first step correct), to  $J$  (all  $J$  steps correct). An IRT model dealing with such ordinal responses is the partial credit model (PCM; Andersen, 1973; Andrich, 1978) or its extension, the generalized partial credit model (GPCM; Muraki, 1992, 1993). It is a mathematically convenient generalization of the dichotomous two-parameter logistic (2PL) model (which is in turn a generalization of the Rasch model) for more than two

\*Correspondence should be addressed to Paul-Christian Bürkner, Fliehdnerstr. 21, 48151 Münster, Germany (email: paul.buerkner@gmail.com).

possible response categories. As such, it combines the ability of an examinee with the item difficulty and discrimination and maps them to the probability of each response category.

Although alternative ordinal models exist and are commonly applied in practice as well – most notably the graded response model (Walker & Duncan, 1967; van der Ark, 2001) and the sequential model (Tutz, 1990, 2011) – the GPCM is probably the most frequently used ordinal model in psychological research. For example, it has been applied in many large-scale educational assessment studies such as PISA, TIMSS, PIRLS, and NAEP (Allen, Donoghue, & Schoeps, 2001; Martin, Gregory, & Stemler, 2000; OECD, 2017). In addition, there is a great body of literature discussing application of the GPCM in large-scale assessment studies from a more technical perspective (see, for example, Mazzeo & Davier, 2014; von Davier & Sinharay, 2010; von Davier & Sinharay, 2014). These studies have mostly been concerned with the comparability of large-scale results as well as model fitting itself, while to our knowledge optimizing the statistical properties of the resulting parameter estimates has not yet been systematically investigated. We believe that, given the GPCM's broad application in practice, it is of great relevance to investigate the conditions under which the GPCM performs best.

Applying principles of optimal design (Atkinson, Donev, & Tobias, 2007; Berger & Wong, 2009; Holling & Schwabe, 2013) to IRT models can lead to considerable efficiency gains and reduce costs of test administration by reducing the number of items and/or subjects required to achieve the same level of precision (Holling & Schwabe, 2016). In IRT, one typically distinguishes between two types of optimal design problem. The first is about selecting items with specific properties for the efficient estimation of person parameters, while the second is about selecting persons with specific abilities for the efficient estimation of item parameters. These problems are referred to as optimal test design and optimal sampling design, respectively (Holling & Schwabe, 2016), with the former being the more common (van der Linden, 2006). Optimal test designs are relevant in so-called calibrated tests, in which item parameters were estimated in prior studies and can thus be considered as known with reasonable precision. The design question is to select the items from the calibrated item pool most appropriate for the participating persons. Optimal test designs are also applicable in tests making use of automatic item generation (see, for example, Geerlings, Glas, & van der Linden, 2011). In this case, the item parameters can be inferred from the set of rules which were used to create the items, provided that the influence of the rules on the item parameters has previously been investigated.

Most IRT models, including the GPCM, are non-linear in the parameters, so that the related optimal designs are dependent on the parameters to be estimated. Such optimal designs that are optimal only for certain parameter values – but not for others – are called locally optimal designs. Among others, locally optimal designs for the GPCM will be investigated in the present paper. Despite being relevant for theoretical reasons and often being a prerequisite more advanced optimal designs, locally optimal designs are of limited practical use themselves as one has to guess the true value of the parameter before constructing the design. If the guess is far away from the truth, such locally optimal design might become relatively inefficient.

There are different possibilities for overcoming the problem of locally optimal designs. One solution considered in the present paper are so-called Bayesian optimal designs. Instead of trying to find an optimal design for a specific parameter value, one specifies a weight distribution over the set of possible parameters to express the uncertainty in the subject or item parameters. For instance, when generating a test design, one might assume that a subject's ability will be somewhere between  $-2$  and  $2$ , with all abilities in between

being equally likely. This would translate into a uniform distribution over the interval  $[-2, 2]$ . A design that is optimal given the specified uncertainty in the parameter values is then called Bayes optimal.

The paper is structured as follows. In Sections 2 and 3 the GPCM is introduced in detail and locally optimal designs are investigated. In Section 4 Bayesian optimal designs are introduced and discussed in relation to the GPCM. Examples for common weight distributions are given in Section 5. The paper is concluded in Section 6 with a discussion of the findings. All proofs are provided in the Appendix.

## 2. The model

In a test situation with ordinal items, the response of person  $p$  on item  $i$  is denoted by  $Y_{pi}$ , which can take on one of  $J + 1$  values ranging from 0 to  $J$ . In the GPCM, the probability  $\pi_{pij}$  that person  $p$  achieves category  $j \in \{0, \dots, J\}$  on item  $i$  is given (cf. Muraki, 1992, 1993) by

$$\pi_{pij} := \pi_j(\theta_p, \boldsymbol{\tau}_i, \boldsymbol{\alpha}_i) := P(Y_{pi} = j; \theta_p, \boldsymbol{\tau}_i, \boldsymbol{\alpha}_i) := \frac{\exp(\sum_{s=1}^j \alpha_{is}(\theta_p - \tau_{is}))}{\sum_{k=0}^J \exp(\sum_{s=1}^k \alpha_{is}(\theta_p - \tau_{is}))}, \quad (1)$$

with

$$\sum_{s=1}^0 \alpha_{is}(\theta_p - \tau_{is}) := 0 \quad (2)$$

for notational convenience. The model results from the assumption that, given only two adjacent categories, the probability for the higher of the two is given by the dichotomous 2PL model,

$$P(Y = j; \theta_p, \boldsymbol{\tau}_i, \boldsymbol{\alpha}_i | Y_{pi} \in \{j, j-1\}) = \frac{\exp(\alpha_{ij}(\theta_p - \tau_{ij}))}{1 + \exp(\alpha_{ij}(\theta_p - \tau_{ij}))}. \quad (3)$$

The parameter  $\theta_p$  denotes the ability of the  $p$ th person. The higher  $\theta_p$ , the higher the probability of reaching higher categories (Agresti, 2010). The parameter vector  $\boldsymbol{\tau}_i = (\tau_{i1}, \dots, \tau_{ij})$  denotes the *thresholds* of the  $i$ th item. Thresholds in the GPCM can be interpreted as follows. If a person has an ability equal to  $\tau_{ij}$ , the probabilities for this person of achieving category  $j$  and  $j - 1$  are equal, that is,  $\pi_{pij} = \pi_{pi(j-1)}$  (i.e. the  $\tau_{ij}$  are the intersection points of the respective category response curves). The higher the thresholds, the lower the probability of achieving higher categories. This does not imply, however, that thresholds must be ordered in the sense that  $\tau_1 \leq \tau_2 \leq \dots \leq \tau_J$  (Adams, Wu, & Wilson, 2012). Rather, all thresholds may take on any real value. The vector  $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{ij})$ , where  $\alpha_{ij} > 0$ , denotes the positive *discrimination* parameters of the  $i$ th item. The higher  $\alpha_{ij}$ , the steeper the curve favouring the  $j$ th category over the  $(j - 1)$ th category with increasing  $\theta_p$ . Note that the GPCM is sometimes defined as having only a single discrimination parameter, which is assumed constant across categories, instead of a vector of discriminations. As the latter is a generalization of the former, all of our results may be applied to GPCMs with a single discrimination parameter as well.

Since the focus of the present paper is on optimal test designs, we consider the item parameters  $\tau_i$  and  $\alpha_i$  to be known (i.e. chosen by the design) and the person parameters  $\theta_p$  as the quantities to be estimated. This situation arises in calibrated tests where item parameters were estimated in prior studies and the design question is to select the most appropriate items from the item pool. Assuming conditional independence of the items for given  $\theta_p$ , the joint density of all items is simply the product of the single-item densities see (for a detailed discussion about conditional independence in IRT models; van der Linden & Glas, 2010). For notational convenience, we will suppress the indices  $p$  and  $i$  in the following where appropriate.

The original PCM was first derived by Rasch (1961) and subsequently by Andersen (1973), Andrich (1978), Masters (1982), and Fischer (1995) each with a different but equivalent formulation (considering the special case of discrimination parameters fixed to 1; cf. Fischer, 1995; Adams *et al.*, 2012). Andersen (1973) and Fischer (1995) derived the PCM in an effort to find a model that allows the independent estimation of person and item parameters – a highly desirable property – for ordinal variables. Andrich (1978, 2005) provided another derivation: When two dichotomous processes are independent, four results can occur: (0,0), (1,0), (0,1), (1,1). Using the Rasch model for each of the two processes, the probability of the combined outcomes is given by the polytomous Rasch model (Andersen, 1973; cf. Wilson, 1992; Wilson & Adams, 1993). When thinking of these processes as steps between ordered categories, (0,0) corresponds to  $Y = 0$ , (1,0) corresponds to  $Y = 1$ , and (1,1) corresponds to  $Y = 2$ . The event (0,1), however, is assumed to be impossible because the second step cannot be successful when the first step was not. For an arbitrary number of ordered categories, Andrich (1978) proved that the polytomous Rasch model becomes the PCM (with discrimination parameters fixed to 1) when considering the set of possible events only. The GPCM, which generalizes the 2PL model to more than two ordinal response categories, was later proposed by Muraki (1992).

### 3. Locally optimal designs

In the context of optimal test designs for the GPCM, an experimental design is the set  $\xi := \{(\tau_1, \alpha_1), (\tau_2, \alpha_2), \dots, (\tau_N, \alpha_N)\}$  of known parameters of the  $N$  administered items. It can be chosen by the experimenter in order to optimize the information of the experiment. Note that each  $\tau_i$  and each  $\alpha_i$  ( $i \in \{1, \dots, N\}$ ) is a vector of length  $J$ . The set of all designs is denoted by  $\Xi$ . When only considering a single item, we simply write  $(\tau, \alpha)$  instead of  $\xi$  and drop the index  $i$ .

There are several optimal design criteria discussed in the literature (see Atkinson *et al.*, 2007, for an overview), but arguably the most common criterion is D-optimality. A design  $\xi$  is called D-optimal if it minimizes the determinant of the estimator's covariance matrix. In the present context, the parameter  $\theta$  to be estimated is the unidimensional person parameter so that the determinant of the covariance matrix reduces to the unidimensional variance. Thus, we can define a D-optimal design as

$$\xi^* := \arg \min_{\xi \in \Xi} \text{Var}(\hat{\theta}; \xi), \quad (4)$$

where  $\hat{\theta}$  is an estimator of the parameter  $\theta$ . According to the Cramér–Rao bound (Cramér, 2016; Rao, 1992), the variance of an unbiased estimator cannot be smaller than the inverse of the Fisher information  $M$ ,

$$\text{Var}(\hat{\theta}, \hat{\xi}) \geq M(\theta, \xi)^{-1}. \quad (5)$$

Thus, for (asymptotically) efficient estimators, we can equivalently maximize the Fisher information instead of minimizing the variance of the estimator. Under certain regularity conditions (cf. Lehmann & Casella, 2006),  $M$  can be written as

$$M(\theta, \xi) := \text{Var} \left[ \frac{d}{d\theta} \log f(Y; \theta, \xi) \right] = -\mathbb{E} \left[ \frac{d^2}{d\theta^2} \log f(Y; \theta, \xi) \right], \quad (6)$$

where  $f(Y; \theta, \xi)$  is the product density of all items given the ability parameter  $\theta$  and the design  $\xi$ , assuming conditional independence of the items. Due to additivity of the Fisher information in this case, we could equivalently write (6) as the sum of the single-item Fisher information values. Most often in optimal design theory, the aim is to minimize the variance of the estimator. For asymptotic efficient estimators (such as maximum likelihood estimators), which by definition meet the Cramér-Rao bound, this is equivalent to maximizing the Fisher information, and we will use this approach in the derivation of optimal designs in the present paper.

**Proposition 1.** *Defining  $A_j := \sum_{s=1}^j \alpha_s$ , the Fisher information  $M(\theta, \tau, \alpha)$  of a single item following the GPCM is given by*

$$M(\theta, \tau, \alpha) = \sum_{j=1}^J A_j^2 \pi_j - \left( \sum_{j=1}^J A_j \pi_j \right)^2. \quad (7)$$

*If we treat  $M$  as a function of the probability vector  $\pi = (\pi_0, \dots, \pi_J)$ , we can obtain the following theorem as an immediate consequence of Proposition 1.*

**Theorem 1.** *The Fisher information  $M$  of a single item treated as a function of the probabilities  $\pi$  has a unique global maximum in  $\{\pi \in \mathbb{R}_+^{J+1} \mid \sum_{j=0}^J \pi_j = 1\}$ , which is given by  $\pi_0^* = \pi_J^* = 1/2$  and  $\pi_1^* = \pi_2^* = \dots = \pi_{J-1}^* = 0$  for given  $A = (A_1, \dots, A_J)$ .*

Theorem 1 implies that an item following the GPCM would be optimal for a person with ability  $\theta_0$ , when the vectors  $\tau$  and  $\alpha$  are chosen so that only the first and last categories have non-zero probability and are equally likely. In other words, such an item is optimal if it is from the dichotomous 2PL model (with scalar discrimination  $\alpha = A_j$ ). Due to additivity of the Fisher information, all items of a locally optimal design  $\xi^*$  for the PCM have to satisfy the above condition.

We want to briefly illustrate the efficiency gain achieved by optimal items with a simple example. Suppose we administer a four-category item with threshold vector  $\tau = (-1, 0, 1)$  and discrimination vector  $\alpha = (1, 1, 1)$ . Then participants with low ( $\theta = -1$ ), medium ( $\theta = 0$ ), and high ( $\theta = 1$ ) ability achieve the categories 0–3 with probabilities  $\pi_l \approx (0.41, 0.41, 0.15, 0.02)$ ,  $\pi_m \approx (0.13, 0.37, 0.37, 0.13)$ , and  $\pi_h \approx (0.02, 0.15, 0.41, 0.41)$ , respectively. Such values are not too uncommon in practice. When compared to an item with optimal thresholds and  $\alpha = (1, 1, 1)$ , the ratio of the Fisher information is  $M^*/M_m \approx 2.86$  for participants with medium ability, and  $M^*/M_l = M^*/M_h \approx 3.75$  for participants with low or high ability. Hence, the variance of

an efficient estimator is about three to four times higher when applying this item rather than optimal items tailored to the participants' abilities.

Unfortunately, there exists no item with a threshold vector  $\tau^*$  that leads to the optimal probabilities of Theorem 1, but it may be approximated:

**Proposition 2.** Define  $\tilde{\tau}_\alpha(c) := (\alpha_J c, 0, 0, \dots, -\alpha_1 c) \in \mathbb{R}^J$  with  $c > 0$  and

$$C_\alpha^J := \left\{ x \in \mathbb{R}^J \mid \forall k \in \{1, \dots, J\} : \alpha_k x_k = -\alpha_{J-k+1} x_{J-k+1} \right\} \quad (8)$$

for some discrimination vector  $\alpha \in \mathbb{R}_+^J$ .

- (a) We have  $\tilde{\tau}_\alpha(c) \in C_\alpha^J$ .
- (b) For each  $\theta \in \mathbb{R}$ , there is some  $\rho = \rho(\theta, \alpha)$  with  $0 < \rho < 1$  such that  $\pi_j(\theta, \tilde{\tau}_\alpha(c), \alpha) \rightarrow \rho$ ,  $\pi_0(\theta, \tilde{\tau}_\alpha(c), \alpha) \rightarrow 1 - \rho$  and  $\pi_j(\theta, \tilde{\tau}_\alpha(c), \alpha) \rightarrow 0$  for  $0 < j < J$  as  $c \rightarrow \infty$ .
- (c) We have  $\rho(0, \alpha) = 1/2$ .
- (d) For  $\tau \in C_\alpha^J$  with symmetric  $\alpha$  satisfying  $\alpha_k = \alpha_{J-k+1}$ , we have that  $\pi_j(\theta, \tau, \alpha) = \pi_{J-j}(-\theta, \tau, \alpha)$  for all  $j \in \{0, \dots, J\}$ .

According to Proposition 2(b) and (c), we can formally define  $\tau^*$  as  $\tau^* := \lim_{c \rightarrow \infty} \tilde{\tau}_\alpha(c)$  when assuming  $\theta_0 = 0$ , which can be achieved by a location shift without loss of generality. Proposition 2 may be generalized to arbitrary  $\theta_0$ , but because of the location shift argument, this is not required for the present paper. The sequence  $\tilde{\tau}_\alpha(c)$  is not the only one that satisfies criteria (a–d) of Proposition 2. However, for the purpose of the present paper, it is completely sufficient that we know at least one such sequence exists. Of course, it is impossible to create items with a threshold vector exactly equal to  $\tau^*$ . Such an optimal item would have infinite and negative infinite first and last thresholds, respectively, while all other thresholds would not be estimable. In other words, for a fixed number of more than two response categories, a locally optimal design does not exist. However, one may choose reasonably large values of  $c$  so that the GPCM closely approximates the dichotomous 2PL model with corresponding locally optimal design. The latter is well known as a one-point design, in which all items have difficulty equal to the person's ability. If additionally the discrimination is treated as part of the design, items should optimally have discrimination as high as possible. This may also be inferred directly from Proposition 1 when considering the special case of only two response categories.

#### 4. Bayesian optimal designs

From a Bayesian perspective, a locally optimal design is the result of setting all prior mass to one point  $\theta_0$ . This seems quite infeasible and is only justified if the true ability is not too far away from  $\theta_0$ , or otherwise the locally optimal design in  $\theta_0$  might perform poorly. One solution is to account for the *a priori* uncertainty in the ability parameter by imposing a non-degenerate weight distribution  $\Pi$  over the parameter space  $\Theta$ . An optimal design  $\xi^* \in \Xi$  taking this uncertainty into account is called a Bayesian optimal design. It is obtained via some sort of averaging over the locally optimal designs' Fisher information, where the Fisher information of each locally optimal design is weighted according to the

weight distribution (Atkinson *et al.*, 2007). In the present paper, we will consider two common Bayesian design criteria that each perform a different kind of averaging (Firth & Hinde, 1997):

$$\psi_0(\xi) := \mathbb{E}[\log(M(\theta, \xi))], \quad (9)$$

$$\psi_{-1}(\xi) := -\mathbb{E}[M(\theta, \xi)^{-1}], \quad (10)$$

where

$$\mathbb{E}[X] = \int X \Pi(\theta) d\theta \quad (11)$$

is the expected value of a random variable  $X$  under the distribution  $\Pi$  of  $\theta$ . The criterion  $\psi_0$  considers the average logarithm of the Fisher information, while the criterion  $\psi_{-1}$  aims to minimize the average asymptotic variance. Of those two,  $\psi_0$  is mathematically more convenient and has a natural Bayesian interpretation (Chaloner & Larntz, 1989). Both criteria are concave with respect to the experimental design  $\xi$  (Firth & Hinde, 1997). Note that, in general, these design criteria are not fully Bayesian in the sense that they do not require computing a posterior distribution. Instead, they borrow the concept of prior distributions to obtain robust designs for frequentist analysis via maximum likelihood methods. Hence, the designs discussed here are sometimes referred to as ‘pseudo-Bayesian’ designs (Firth & Hinde, 1997). For a review of fully Bayesian design criteria, see Chaloner and Verdinelli (1995).

In theory, the weight distribution  $\Pi$  of  $\theta$  may take any form. In practice, however, one typically uses symmetric weights, both for mathematical convenience and because there is rarely any prior information available that favours an asymmetric weight distribution for an unbounded parameter such as the ability  $\theta$ . Accordingly, we will focus on symmetric weight distributions in the following. Within this class, there are (among others) the continuous uniform, normal, and logistic distributions, each forming a so-called location–scale family of distributions.

For the Rasch model, Bayesian optimal designs have already been investigated in Graßhoff, Holling, and Schwabe (2012), and the aim of the present section aims is to generalize these results to the GPCM. The following lemma underlines the relevance of locally optimal designs in the context of Bayesian optimal design criteria.

**Lemma 1.** *If the weight distribution  $\Pi$  is symmetric around some ability  $\theta_0$ ,  $\alpha$  is fixed to any vector in  $\mathbb{R}_+^J$ , and only  $\tau$  is considered variable, the Bayes optimal one-point design with respect to  $\psi_0$  and  $\psi_{-1}$  is the locally optimal design for  $\theta_0$ , that is,  $\tau = \tau^*$ .*

According to Lemma 1, the same problem that occurs for the locally optimal design occurs for the optimal one-point design under the Bayesian criteria as well. That is, it does not exist for a fixed number of more than two response categories, and when considering the number of categories variable, the GPCM reduces to the dichotomous 2PL model.

Accordingly, we need to investigate the optimal discrimination parameter  $\alpha$  in the case of only two response categories.

**Lemma 2.** *If the weight distribution  $\Pi$  is symmetric around some ability  $\theta_0$ ,  $\tau$  is fixed to  $\tau^*$ , and  $\alpha \in \mathbb{R}_+$  is considered variable, then:*

(a)  $\psi_0$  is maximized by the unique solution  $\alpha_0^*$  of

$$\alpha \int \theta \pi_1(\theta, \tau^*, \alpha) \Pi(\theta) d\theta = 1; \quad (12)$$

(b)  $\psi_{-1}$  is maximized by the unique solution  $\alpha_{-1}^*$  of

$$\alpha \int \frac{\theta}{\pi_0(\theta, \tau^*, \alpha)} \Pi(\theta) d\theta = \int \frac{1}{\pi_1(\theta, \tau^*, \alpha) \pi_0(\theta, \tau^*, \alpha)} \Pi(\theta) d\theta. \quad (13)$$

Lemma 2 is in line with our intuition that the optimal  $\alpha$  decreases with increasing scale of  $\Pi$ . Lemmas 1 and 2 do not state under what conditions the optimal one-point design is Bayes optimal. To answer this question, we have to make use of approximate design theory (cf. Kiefer, 1974) and introduce *sensitivity functions* measuring the quality of a design (cf. Graßhoff et al., 2012). If the sensitivity function is uniformly bounded at a certain value for a given design  $\xi^*$ , no improvement is possible and  $\xi^*$  will be optimal. For non-linear models, sensitivity functions of D-optimal designs are given in Firth and Hinde (1997). For the GPCM, these sensitivity functions can be written as

$$\phi_v(\tau, \alpha, \xi) = \frac{E[M(\xi, \theta)^v M(\xi, \theta)^{-1} M(\tau, \alpha, \theta)]}{E[M(\xi, \theta)^v]} - 1, \quad (14)$$

where  $v \in \{-1, 0\}$  corresponds to the two Bayesian design criteria discussed above. Since, for the GPCM, the design criteria are concave, Bayes optimality of a design  $\xi^*$  is equivalent to

$$\sup_{(\tau, \alpha) \in \mathcal{X}} \phi_v(\tau, \alpha, \xi^*) = 0, \quad (15)$$

where  $\mathcal{X}$  denotes the design region of a single item. Moreover, when  $\xi^*$  is optimal,  $\phi_v(\tau, \alpha, \xi^*) = 0$  holds if and only if  $(\tau, \alpha)$  is a design point of  $\xi^*$ . It may now be asked under what conditions the optimal one-point design is also Bayes optimal.

**Theorem 2.** *Necessary conditions for the threshold  $\tau^*$  to be Bayes optimal under a symmetric weight distribution  $\Pi$  and fixed discrimination  $\alpha$  are*

$$\psi_0 : \int \pi_J(\theta, \tau^*, \alpha) \pi_0(\theta, \tau^*, \alpha) \Pi(\theta) d\theta \geq \frac{1}{6}, \quad (16)$$



$$\psi_{-1} : \int (\pi_J(\theta, \tau^*, \alpha) \pi_0(\theta, \tau^*, \alpha))^{-1} \Pi(\theta) d\theta \leq 6. \quad (17)$$

For a fixed discrimination of  $\alpha = 1$ , Theorem 2 provides the same conditions for the GPCM as can be found for the Rasch model (see Graßhoff et al., 2012). By Jensen's inequality, the necessary condition for  $\psi_{-1}$  implies the condition for  $\psi_0$ , that is, the former condition is more restrictive. For scale families of symmetric weight distributions, Theorem 2 may be further extended.

**Theorem 3.** *Let  $(\Pi_s)_{s>0}$  be a scale family of symmetric distributions around some ability  $\theta_0$ .*

- (a) *There exist unique values  $s_{-1}$  and  $s_0$ , with  $0 < s_{-1} \leq s_0 \leq \infty$ , such that the necessary conditions of Theorem 2 on  $\psi_{-1}$  and  $\psi_0$  are satisfied for all  $s \leq s_{-1}$  and  $s \leq s_0$ , respectively, and violated otherwise.*
- (b) *There exists a unique threshold  $\tilde{s}_0$ ,  $0 \leq \tilde{s}_0 < s_0$ , such that the optimal one-point design  $\tau^*$  is Bayes optimal with respect to  $\psi_0$  for all weight distributions  $\Pi$  with  $s \leq \tilde{s}_0$  and fails to be optimal for  $s > \tilde{s}_0$ .*

According to Theorem 3, the optimal one-point design for a symmetric scale distribution remains Bayes optimal as long as the scale parameter does not exceed a certain value that depends on the design criterion, the (now scalar) discrimination  $\alpha$ , and on the location–scale family. As  $\int \pi_J \pi_0 \Pi(\theta) d\theta$  decreases monotonically with increasing  $\alpha$ , the values  $\tilde{s}_0$ ,  $s_0$ , and  $s_{-1}$  decrease monotonically with increasing  $\alpha$  as well. For some common distribution families, examples are discussed in more detail in the next section.

Finally, we want to turn our attention to Bayes optimal designs for the dichotomous 2PL model, when both  $\tau$  and  $\alpha$  are allowed to vary. It is of particular interest to determine in which cases the optimal one-point designs derived in Lemma 2 are also Bayes optimal. In the next section we show numerically that for certain important weight distributions under  $\phi_0$  or  $\phi_{-1}$ , the optimal one-point design is never Bayes optimal. This implies that the Bayes optimal designs under  $\phi_0$  or  $\phi_{-1}$  (if they exist) have to have at least two distinct design points. Whether this holds in general for all symmetric weight distributions requires further investigation.

## 5. Examples

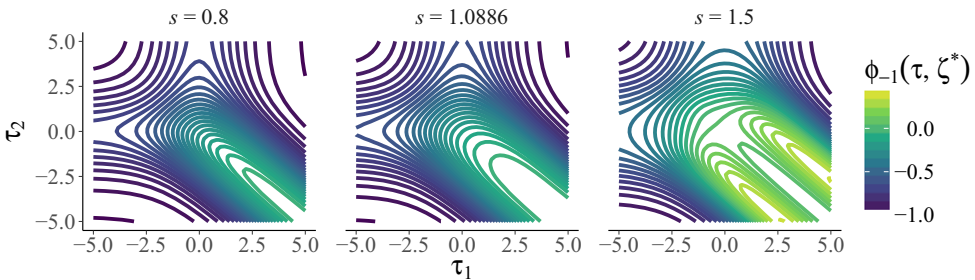
Graßhoff et al. (2012), among others, discuss Bayes optimal designs for the Rasch model for the continuous uniform, normal, and logistic families. They also derive the family-specific values of  $s_{-1}$  and  $s_0$  (denoted  $\tau_1$  and  $\tau_2$ , respectively, in their paper) for the Rasch model (i.e. for discrimination fixed to 1). We have already established that, when using the optimal one-point design, the GPCM is equivalent to the 2PL model. We can obtain  $s_0$  and  $s_{-1}$  for the GPCM by dividing the respective values for the Rasch model by the sum  $A_J$  of all discrimination parameters. Theoretically, this allows us to numerically investigate whether the necessary conditions on the optimality of the one-point design are also sufficient, by computing the sensitivity function across the design space. In practice, this becomes increasingly complicated for items with more response categories, since the design space is of dimension  $J$  (or  $2J$  when also varying  $\alpha$ ). Accordingly, we will focus on items with up to three response categories, which may also be visualized conveniently.

For each example, we will consider two cases: (1) three response categories with varying  $\tau$  and discrimination parameters fixed to 1 (hence  $A_j = j$  for all  $j$ ); (2) two response categories with varying  $\tau$  and  $\alpha$ .

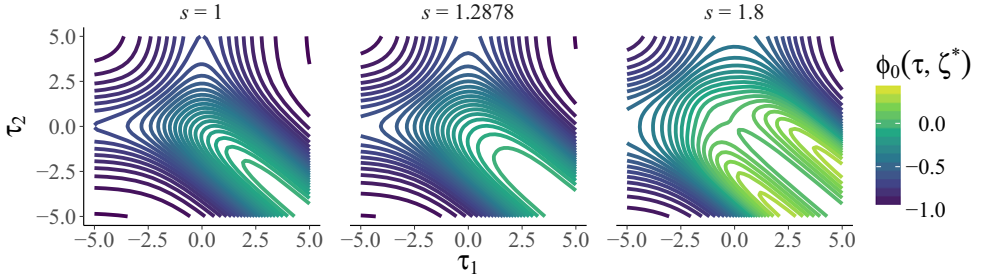
**Example 1.** (Continuous uniform distribution) For the continuous uniform distribution with support in  $[\theta_1, \theta_2] = [\theta_0 - s, \theta_0 + s]$  and constant density  $f(\theta) = \frac{1}{2s}$ , we have  $s_{-1} \approx 2.1773/J$  and  $s_0 \approx 2.5757/J$ . The scalars 2.1773 and 2.5757 are numeric solutions of the equations  $\exp(s) - \exp(-s) = 4s$  and  $(3-s) \exp(s) = 3 + s$ , respectively (Graßhoff et al., 2012). For  $J = 2$ , the sensitivity function is displayed in Figures 1 and 2, with  $s_{-1}$  and  $s_0$  being displayed in the center. In the figures, yellowish regions are values of the sensitivity function that are greater than zero, indicating non-optimality of the one-point design. At  $s_{-1}$  and  $s_0$ , or smaller values, the sensitivity functions do not exceed zero, and are unimodal with maximal value zero at  $\tau^*$ . This demonstrates, at least numerically and for  $J = 2$ , that the necessary conditions of Theorem 2 are also sufficient for the continuous uniform weight distribution. When  $s$  is further increased, the one-point design becomes visibly non-optimal and a two-point design appears to be favoured.

For the dichotomous 2PL model with varying  $\tau$  and  $\alpha$ , we have  $\alpha_{-1}^* \approx 3.1560$  for  $s = s_{-1} \approx 2.1773$ . However, the necessary condition (17) for  $\phi_{-1}$  is only satisfied for  $\alpha \leq \alpha_{-1}^+ \approx 2$ , implying that the optimal one-point design  $(\tau^*, \alpha_{-1}^*)$  is not Bayes optimal (see also Figure 3 (left)). The same result may be obtained for other values of the scale  $s$  in the sense that  $\alpha_{-1}^*$  exceeds  $\alpha_{-1}^+$  independently of  $s$ . Similarly, we have  $\alpha_0^* \approx 3.6186$  for  $s = s_0 \approx 2.5757$ , but the necessary condition (16) for  $\phi_0$  is only satisfied for  $\alpha \leq \alpha_0^+ \approx 2$ , (see also Figure 3 (right)). Again,  $\alpha_0^* > \alpha_0^+$  can be shown to hold for other values of  $s$  as well.

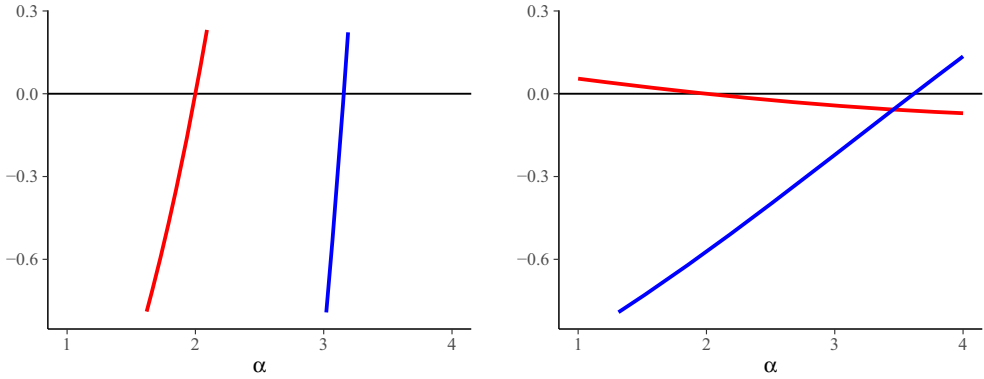
**Example 2.** (Normal distribution) Similar results are obtained for the normal weight distribution. Here, the scale  $s$  is simply the standard deviation parameter of the normal distribution. The critical values are approximately  $s_{-1} \approx 1.177/J$  and  $s_0 \approx 1.683/J$ . Again, numerical computation shows that the sensitivity functions do not exceed zero for  $s \leq s_{-1}$  or  $s \leq s_0$  (see Figures 4 and 5). Thus, the necessary conditions appear to be sufficient for the normal weight distribution when  $J = 2$ . Accordingly, the optimal one-point design remains Bayes optimal for items with three response categories as long as the standard deviation does not exceed the critical values.



**Figure 1.** Contour of the sensitivity function for  $\psi_{-1}$  for three response categories for a uniform weight distribution with varying values of the scale parameter  $s$ .



**Figure 2.** Contour of the sensitivity function for  $\psi_0$  for three response categories for a uniform weight distribution with varying values of the scale parameter  $s$ .

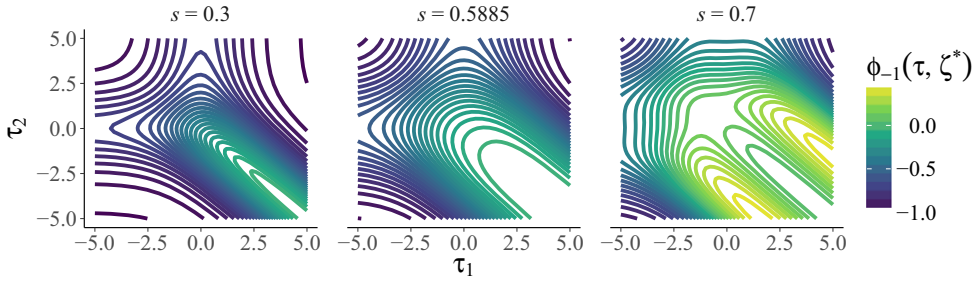


**Figure 3.** Illustration of the optimal one-point discrimination parameter of the 2PL model in relation to the necessary condition for Bayes optimality. Left: criterion  $\phi_{-1}$  with a uniform weight distribution with scale  $s = 2.1773$ . Right: criterion  $\phi_0$  with a uniform weight distribution with scale  $s = 2.5757$ . The intersection of the blue and black lines indicates  $\alpha_{-1}^*$  and  $\alpha_0^*$ , respectively, while the intersection of the red and black lines indicates  $\alpha_{-1}^+$  and  $\alpha_0^+$ , respectively. Note that  $\alpha_{-1}^* > \alpha_{-1}^+$  and  $\alpha_0^* > \alpha_0^+$ .

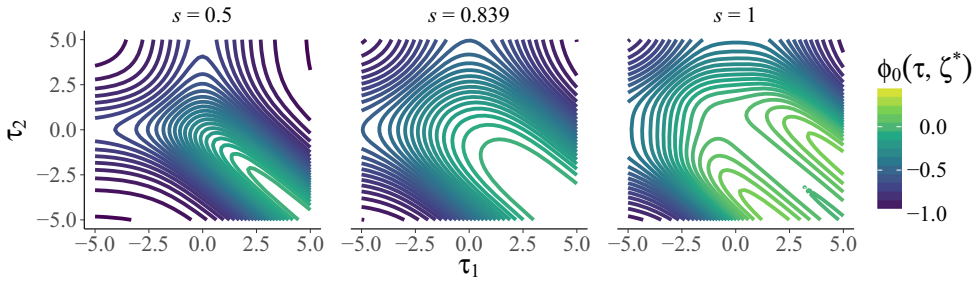
For the dichotomous 2PL model with varying  $\tau$  and  $\alpha$ , we have  $\alpha_{-1}^* \approx 1.3586$  for  $s = s_{-1} \approx 1.177$ , but the necessary condition (17) for  $\phi_{-1}$  is only satisfied for  $\alpha \leq \alpha_{-1}^+ \approx 1$ . Similarly, we have  $\alpha_0^* \approx 1.7350$  for  $s = s_0 \approx 1.683$ , but the necessary condition (16) for  $\phi_0$  is only satisfied for  $\alpha \leq \alpha_0^+ \approx 1.002$ . This may be shown to hold for other values of  $s$  as well.

**Example 3.** (Logistic distribution) For the logistic distribution with scale parameter  $s$ , the critical values are given by  $s_{-1} \approx 0.603/J$  and  $s_0 = 1/J$ . According to Figure 6, the necessary condition for  $\psi_0$  is also sufficient if  $J = 2$ . However, the sensitivity function of  $\psi_{-1}$  behaves somewhat differently for the logistic weight distribution in the sense that the necessary condition is apparently not sufficient (see Figure 7 [centre]). However, for scale values only slightly below  $s_{-1}$ , the optimal one-point design appears to be Bayes optimal (see Figure 7 [left]), indicating the best possible sufficient condition, very close to the necessary condition derived in the present paper.

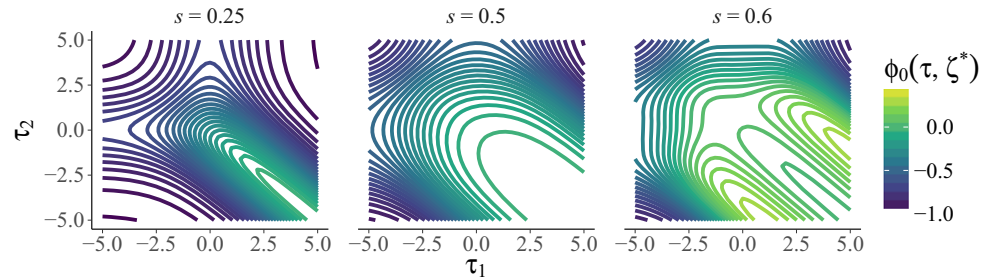
For the dichotomous 2PL model with varying  $\tau$  and  $\alpha$ , we have  $\alpha_{-1}^* \approx 2.6518$  for  $s = s_{-1} \approx 0.603$ , but the necessary condition (17) for  $\phi_{-1}$  is only satisfied for



**Figure 4.** Contour of the sensitivity function for  $\psi_{-1}$  for three response categories for a normal weight distribution with varying values of the scale parameter  $s$ .



**Figure 5.** Contour of the sensitivity function for  $\psi_0$  for three response categories for a normal weight distribution with varying values of the scale parameter  $s$ .

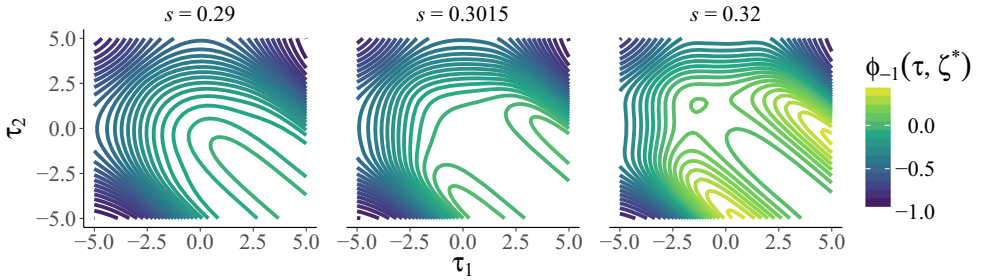


**Figure 6.** Contour of the sensitivity function for  $\psi_0$  for three response categories for a logistic weight distribution with varying values of the scale parameter  $s$ .

$\alpha \leq \alpha_{-1}^+ \approx 1.953$ . Similarly, we have  $\alpha_0^* \approx 2.9217$  for  $s = s_0 \approx 1$ , but the necessary condition (16) for  $\phi_0$  is only satisfied for  $\alpha \leq \alpha_0^+ \approx 1.6868$ . This may be shown to hold for other values of  $s$  as well.

## 6. Discussion

In the present paper we investigated locally optimal designs as well as Bayes optimal designs for the GPCM. We demonstrated that for a fixed number of more than two response categories, the locally optimal design does not exist. However, when



**Figure 7.** Contour of the sensitivity function for  $\psi_{-1}$  for three response categories for a logistic weight distribution with varying values of the scale parameter  $s$ .

allowing the number of categories to vary, the locally optimal design exists and assigns non-zero probability only to the first and last response categories independently of a person's ability. That is, when using such a design, the GPCM reduces to the ordinary 2PL model. This perhaps paradoxical result will be discussed in more detail later on.

When fixing the discrimination parameters, the locally optimal one-point design proved to be relatively robust. In particular, when allowing the number of response categories to vary, it turned out to be Bayes optimal under the two most common Bayesian design criteria and symmetric weight distributions with not too large variation. It must be emphasized that the maximum variation such that it remained Bayes optimal for both criteria can be considered as sufficiently large for applications. For instance, when using a normal weight distribution with standard deviation of  $s \approx 0.59$  as well as items with three response categories and a discrimination of 1, subjects in the lowest 2.5% ability region (according to the weight distribution) completely solve the task only with probability 9% or less, while subjects in the highest 2.5% ability region completely solve the task with probability 91% or more. Thus, even for substantial variation in subjects' abilities, one-point designs remain comparatively viable.

When focusing on dichotomous responses and thus on the 2PL model, we provided equations whose unique solutions are the optimal discrimination parameters in the set of one-point designs when applying symmetric weight distributions. However, for some important examples, we established numerically that these discrimination parameters do not lead to Bayes optimal designs. That is, any Bayes optimal design for the 2PL model with varying threshold and discrimination (if existent) must have a minimum of two distinct design points at least for the examples discussed in the present paper.

Despite being mathematically reasonable, our results reveal practical problems when trying to design items with optimal properties. An item with more than two response categories, on which subjects score on the first or last categories (nearly) always – independent of their ability – may be practically impossible to construct. One could only aim to reduce the probability of intermediate responses to approximate the optimal design. However, this would contradict the very idea of using ordinal items instead of dichotomous ones. The obvious solution, to completely remove partial credit from the items and score only as true or false, might be tempting, but will definitely lead to information loss – and thus less efficient designs – for complicated items consisting of multiple tasks. This does not invalidate the GPCM in general, as the lack of a reasonable

locally optimal design does not affect other mathematical or psychometric properties of the model. Further, the Fisher information or other criterion may still be used to compare items or to choose items with the highest information among those being available.

Still, it is of great interest to study optimal designs of alternative ordinal models such as the graded response model (van der Ark, 2001; Walker & Duncan, 1967) or the sequential model (Tutz, 1990, 2011), which potentially favour items with properties much closer to real items. If so, one should also evaluate their potential use in large-scale educational assessment studies such as PISA, which previously applied the GPCM.

For future research, some open questions remain. It is still unclear how the optimal designs derived here perform for asymmetric weight distributions due to the complexity of the derivatives and integrals involved. Furthermore, it must be noted that optimal test designs and optimal sampling designs are not equivalent as there are multiple threshold as well as discrimination parameters per item, but only one ability parameter per person. Accordingly, derivation of optimal sampling designs for the GPCM still remains an open topic. Finally, as discussed above, there are other reasonable ordinal models, whose optimal designs are potentially of great relevance in practice. While locally optimal designs for the graded response model have been derived in (Schmidt and Schwabe (2015), Bayes optimal designs for this model as well as optimal designs in general for the sequential model remain to be investigated.

## References

- Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, 72(4), 547–573. <https://doi.org/10.1177/0013164411432166>
- Agresti, A. (2010). *Analysis of ordinal categorical data*. Chichester, UK: John Wiley & Sons. <https://doi.org/10.1002/9780470594001>
- Allen, N., Donoghue, J., & Schoeps, T. (2001). *The NAEP 1998 technical report (NCES 2001509)*. Washington, DC: National Center for Education Statistics.
- Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26(1), 31–44. <https://doi.org/10.1111/j.2044-8317.1973.tb00504.x>
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. <https://doi.org/10.1007/BF02293814>
- Andrich, D. (2005). The Rasch model explained. In *Applied Rasch measurement: A book of exemplars* (pp. 27–59). Berlin, Germany: Springer. <https://doi.org/10.1007/1-4020-3076-2>
- Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs, with SAS* (Vol. 34). Oxford, UK: Oxford University Press. [https://doi.org/10.1007/978-3-642-04898-2\\_434](https://doi.org/10.1007/978-3-642-04898-2_434)
- Berger, M. P., & Wong, W.-K. (2009). *An introduction to optimal designs for social and biomedical research* (Vol. 83). Chichester, UK: John Wiley & Sons. <https://doi.org/10.1002/9780470746912>
- Chaloner, K., & Larntz, K. (1989). Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21(2), 191–208. [https://doi.org/10.1016/0378-3758\(89\)90004-9](https://doi.org/10.1016/0378-3758(89)90004-9)
- Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10, 273–304. <https://doi.org/10.1214/ss/1177009939>
- Cramér, H. (2016). *Mathematical methods of statistics* (Vol. 9). Princeton, NJ: Princeton University Press.
- Firth, D., & Hinde, J. (1997). On Bayesian D-optimum design criteria and the equivalence theorem in non-linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 793–797. <https://doi.org/10.1111/1467-9868.00096>

- Fischer, G. H. (1995). The derivation of polytomous Rasch models. In *Rasch models* (pp. 293–305). Berlin, Germany: Springer. <https://doi.org/10.1007/978-1-4612-4230-7>
- Geerlings, H., Glas, C. A., & van der Linden, W. J. (2011). Modeling rule-based item generation. *Psychometrika*, *76*(2), 337–359. <https://doi.org/10.1007/s11336-011-9204-x>
- Graßhoff, U., Holling, H., & Schwabe, R. (2012). Optimal designs for the Rasch model. *Psychometrika*, *77*, 710–723. <https://doi.org/10.1007/s11336-012-9276-2>
- Holling, H., & Schwabe, R. (2013). An introduction to optimal design: Some basic issues using examples from dyscalculia research. *Zeitschrift für Psychologie [Journal of Psychology]*, *221*, 124–144. <https://doi.org/10.1027/2151-2604/a000142>
- Holling, H., & Schwabe, R. (2016). Statistical optimal design theory. In *Handbook of item response theory, volume two: Statistical tools* (pp. 313–340). London, UK: Chapman and Hall/CRC.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, *2*, 849–879. <https://doi.org/10.1214/aos/1176342810>
- Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Berlin, Germany: Springer Science & Business Media. <https://doi.org/10.1007/978-1-4615-8071-3>
- Martin, M. O., Gregory, K. D., & Stemler, S. E. (2000). *TIMSS 1999 technical report*. Chestnut Hill, MA: International Study Center.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Mazzeo, J., & von Davier, M. (2014). *Linking scales in international large-scale assessments*. Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis, 228–258.
- Muraki, E. (1992). A generalized partial credit model: Application of an em algorithm. *ETS Research Report Series*, *1992*(1), 1–35. <https://doi.org/10.1002/j.2333-8504.1992.tb01436.x>
- Muraki, E. (1993). Information functions of the generalized partial credit model. *ETS Research Report Series*, *1993*(1), 1–32. <https://doi.org/10.1002/j.2333-8504.1993.tb01538.x>
- OECD (2017). *PISA 2015: Technical report*. Retrieved from <http://www.oecd.org/pisa/data/2015-technical-report/>
- Peajcariac, J. E., & Tong, Y. (1992). *Convex functions, partial orderings, and statistical applications*. Cambridge, MA: Academic Press.
- Popoviciu, T. (1965). Sur les 'equations alg'ebriques ayant toutes leurs racines réelles [About algebraic equations with only real solutions]. *Mathematica [Mathematics]*, *9*, 129–145.
- Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics* (pp. 235–247). Berlin, Germany: Springer. <https://doi.org/10.1007/978-1-4612-0919-5>
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321–333). Oakland, CA: University of California Press.
- Schmidt, M., & Schwabe, R. (2015). Optimal cutpoints for random observations. *Statistics*, *49*, 1366–1381. <https://doi.org/10.1080/02331888.2015.1006225>
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*(1), 39–55. <https://doi.org/10.1111/j.2044-8317.1990.tb00925.x>
- Tutz, G. (2011). *Regression for categorical data (Vol. 34)*. Cambridge, UK: Cambridge University Press.
- van der Ark, L. A. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*, *25*(3), 273–282. <https://doi.org/10.1177/01466210122032073>
- van der Linden, W. J. (2006). *Linear models for optimal test design*. Berlin, Germany: Springer Science & Business Media. <https://doi.org/10.1007/0-387-29054-0>
- van der Linden, W. J., & Glas, C. A. (2010). Statistical tests of conditional independence between responses and/or response times on test items. *Psychometrika*, *75*(1), 120–139. <https://doi.org/10.1007/s11336-009-9129-9>

- von Davier, M., & Sinharay, S. (2010). Stochastic approximation methods for latent regression item response models. *Journal of Educational and Behavioral Statistics*, 35(2), 174–193. <https://doi.org/10.3102/1076998609346970>
- Von Davier, M., & Sinharay, S. (2014). *Analytics in international large-scale assessments: Item response theory and population models*. Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis, 155–174.
- Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54(1–2), 167–179. <https://doi.org/10.2307/2333860>
- Wilson, M. (1992). The ordered partition model: An extension of the partial credit model. *Applied Psychological Measurement*, 16(4), 309–325. <https://doi.org/10.1177/014662169201600401>
- Wilson, M., & Adams, R. J. (1993). Marginal maximum likelihood estimation for the ordered partition model. *Journal of Educational and Behavioral Statistics*, 18(1), 69–90. <https://doi.org/10.2307/1165183>

Received 9 November 2017; revised version received 28 July 2018

## Appendix

The indices  $p$  and  $i$  for persons and items are suppressed for notational convenience.

*Proof.* (Proposition 1) We abbreviate the normalizing function in the denominator of (1) as  $z (= 1/\pi_0)$  for convenience. For a single item, the response  $Y \in \{0, \dots, J\}$  is categorically distributed so that the log-likelihood  $l$  equals

$$\begin{aligned}
 l(Y; \theta, \boldsymbol{\tau}, \boldsymbol{\alpha}) &:= \log(f(Y; \theta, \boldsymbol{\tau}, \boldsymbol{\alpha})) = \sum_{j=0}^J \mathbf{1}_j(Y) \log(\pi_j) + q \\
 &= \sum_{j=1}^J \mathbf{1}_j(Y) \log(\pi_j) + \left(1 - \sum_{j=1}^J \mathbf{1}_j(Y)\right) \log(\pi_0) + q \\
 &= \sum_{j=1}^J \left(\mathbf{1}_j(Y) \log\left(\frac{\pi_j}{\pi_0}\right)\right) + \log(\pi_0) + q \\
 &= \sum_{j=1}^J \left(\mathbf{1}_j(Y) \sum_{s=1}^j \alpha_s(\theta - \tau_s)\right) - \log(z) + q, \quad (18)
 \end{aligned}$$

where  $f$  is the density of  $Y$  defined in (1),  $\mathbf{1}_j$  is the indicator function for  $Y = j$  and  $q$  is a constant independent of  $\theta$ . The first derivative of  $l$  with respect to  $\theta$  equals

$$\begin{aligned}
 \frac{dl(\theta, \boldsymbol{\tau}, \boldsymbol{\alpha})}{d\theta} &= \sum_{j=1}^J A_j \mathbf{1}_j(Y) - \frac{\sum_{j=1}^J A_j \exp(\sum_{s=1}^j \alpha_s(\theta - \tau_s))}{z} \\
 &= \sum_{j=1}^J A_j (\mathbf{1}_j(Y) - \pi_j). \quad (19)
 \end{aligned}$$

Thus,  $M$  is equal to the variance of the random variable



$$X := \sum_{j=1}^J A_j 1_j(Y). \quad (20)$$

Since  $(1_j(Y))^2 = 1_j(Y)$  and  $1_j(Y) 1_k(Y) = 0$  for  $j \neq k$ , as well as  $\mathbb{E}(1_j(Y)) = \pi_j$ , we have

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{j=1}^J A_j^2 \pi_j - \left( \sum_{j=1}^J A_j \pi_j \right)^2,$$

which completes the proof.  $\square$

*Proof.* (Theorem 1) The random variable  $X$ , defined in (20), is bounded between 0 and  $A_J$ . It is minimal if  $Y = 0$  and maximal if  $Y = J$ . According to Popoviciu's inequality for variances (Popoviciu, 1965; Peajcariac & Tong, 1992),  $\text{Var}(X)$  (and hence  $M$ ) is maximal if and only if  $\pi_0 = \pi_J = 1/2$  and  $\pi_j = 0$  for  $0 < j < J$ .  $\square$

*Proof.* (Proposition 2)

(a) This can easily be checked.

(b) We have

$$\sum_{k=0}^J \exp\left(\sum_{s=1}^k \alpha_s(\theta - \tilde{\tau}_{\alpha s}(c))\right) = 1 + \sum_{k=1}^{J-1} \exp(A_k \theta - \alpha_1 \alpha_j c) + \exp(A_J \theta). \quad (21)$$

Hence, for  $c \rightarrow \infty$ , we have that  $\pi_0 = 1/(1 + \exp(A_J \theta))$ ,  $\pi_j = \exp(A_j \theta)/(1 + \exp(A_J \theta))$  and  $\pi_j = 0$  for all  $1 \leq j < J$ .

(c) This follows directly from the proof of (b).

(d) For  $\tau \in C_\alpha^J$  with  $\alpha$  satisfying  $\alpha_k = \alpha_{J-k+1}$  ( $k \in \{1, \dots, J\}$ ) we have

$$\begin{aligned} \exp\left(\sum_{s=1}^{J-j} \alpha_s(-\theta - \tau_s)\right) &= \exp\left(-A_{J-j}\theta - \sum_{s=1}^{J-j} \alpha_s \tau_s\right) \\ &= \exp\left(-(A_J - A_j)\theta - \sum_{s=1}^j \alpha_s \tau_s\right) \\ &= \exp(-A_j \theta) \exp\left(\sum_{s=1}^j \alpha_s(\theta - \tau_s)\right) \end{aligned} \quad (22)$$

since  $A_J = A_j + A_{J-j}$  and  $\sum_{s=1}^{J-j} \alpha_s \tau_s = \sum_{s=1}^j \alpha_s \tau_s$  for all  $j \in \{0, \dots, J\}$ . Again for  $\tau \in C_\alpha^J$  the denominator of  $\pi_{J-j}(-\theta, \tau, \alpha)$  can be written as

$$\begin{aligned}
\sum_{k=0}^J \exp\left(\sum_{s=1}^k \alpha_s(-\theta - \tau_s)\right) &= \exp(-A_J\theta) \sum_{k=0}^J \exp\left((A_J - A_k)\theta - \sum_{s=1}^k \alpha_s \tau_s\right) \\
&= \exp(-A_J\theta) \sum_{k=0}^J \exp\left(A_{J-k}\theta - \sum_{s=1}^{J-k} \alpha_s \tau_s\right) \\
&= \exp(-A_J\theta) \sum_{k=0}^J \exp\left(\sum_{s=1}^k \alpha_s(\theta - \tau_s)\right).
\end{aligned} \tag{23}$$

Hence, we find that

$$\pi_{J-j}(-\theta, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \frac{\exp(-A_J\theta)}{\exp(-A_J\theta)} \pi_j(\theta, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \pi_j(\theta, \boldsymbol{\tau}, \boldsymbol{\alpha}). \tag{24}$$

□

In the following proofs, we will suppress the dependency of  $M$ ,  $\pi_j$ , and  $\Pi$  on  $\theta$  for notational convenience. Whenever we apply the degenerate locally optimal design  $\tau^*$ , we implicitly reduce the number of categories to two, as in this case only two response categories have positive probability. This way, the derived results remain valid without writing down any limits of the form  $\lim_{\tau \rightarrow \tau^*}$ , thus simplifying the notation.

*Proof.* (Lemma 1) For  $j \geq i$ , the derivatives of  $\pi_j$  with respect to  $\tau_i$  and  $\alpha_i$  equal

$$\begin{aligned}
\frac{d\pi_j}{d\tau_i} &= \alpha_i \left( \exp\left(\sum_{s=1}^j \alpha_s(\theta - \tau_s)\right) \left( \sum_{k=i}^J \exp\left(\sum_{s=1}^k \alpha_s(\theta - \tau_s)\right) - z \right) \right) / z^2 \\
&= \alpha_i \pi_j \left( \sum_{k=i}^J \pi_k - 1 \right) = \alpha_i \left( \pi_j \left( \sum_{k=i}^J \pi_k \right) - \pi_j \right).
\end{aligned} \tag{25}$$

For  $j < i$ , the derivatives equal

$$\begin{aligned}
\frac{d\pi_j}{d\tau_i} &= \alpha_i \left( \exp\left(\sum_{s=1}^j \alpha_s(\theta - \tau_s)\right) \left( \sum_{k=i}^J \exp\left(\sum_{s=1}^k \alpha_s(\theta - \tau_s)\right) \right) \right) / z^2 \\
&= \alpha_i \pi_j \left( \sum_{k=i}^J \pi_k \right).
\end{aligned} \tag{26}$$

Thus, we find

$$\sum_{j=1}^J \frac{d\pi_j}{d\tau_i} = \alpha_i \left( \left( \sum_{j=1}^J \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J \pi_j \right) \tag{27}$$

and hence

$$\begin{aligned}
\frac{dM}{d\tau_i} &= \sum_{j=1}^J A_j^2 \frac{d\pi_j}{d\tau_i} - 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{j=1}^J A_j \frac{d\pi_j}{d\tau_i} \right) \\
&= \alpha_i \left( \left( \sum_{j=1}^J A_j^2 \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J A_j^2 \pi_j \right. \\
&\quad \left. - 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J A_j \pi_j \right) \right).
\end{aligned} \tag{28}$$

Using Proposition 2(b), (28) evaluated at point  $\tau^*$  is equal to

$$\frac{dM}{d\tau_i}(\tau^*, \alpha) = \alpha_i A_j^2 \pi_j \pi_0 (\pi_j - \pi_0). \tag{29}$$

Under suitable regularity conditions on  $\Pi$ , differentiation with respect to  $\tau_i$  and integration with respect to  $\theta$  may be interchanged, so that we arrive at

$$\frac{d\psi_0}{d\tau_i}(\tau^*, \alpha) = \int \frac{dM}{d\tau_i}(\tau^*, \alpha) M(\tau^*, \alpha)^{-1} \Pi d\theta = \alpha_i \int (\pi_j - \pi_0) \Pi d\theta, \tag{30}$$

$$\frac{d\psi_{-1}}{d\tau_i}(\tau^*, \alpha) = \int \frac{dM}{d\tau_i}(\tau^*, \alpha) M(\tau^*, \alpha)^{-2} \Pi d\theta = \frac{\alpha_i}{A_j^2} \int \frac{(\pi_j - \pi_0)}{\pi_j \pi_0} \Pi d\theta, \tag{31}$$

as

$$M(\tau^*, \alpha) = A_k^2 \pi_j \pi_0. \tag{32}$$

By symmetry of  $\Pi$  and  $\pi_i(\cdot, \tau^*)$  (in the sense of Proposition 2(d)) we have

$$\int \pi_j^2 \pi_0 \Pi d\theta = \int \pi_j \pi_0^2 \Pi d\theta, \tag{33}$$

and hence

$$\frac{d\psi_0}{d\tau_i}(\tau^*, \alpha) = \frac{d\psi_{-1}}{d\tau_i}(\tau^*, \alpha) = 0$$

for all  $i \leq J$  due to symmetry. Since  $\psi_v$  ( $v \in \{-1, 0\}$ ) is concave (Firth & Hinde, 1997),  $\tau^*$  is a global maximum of the two design criteria within the set of one-point designs independent of the choice of  $\alpha$ .  $\square$

*Proof.* (Lemma 2) Although we only need  $J=1$  for the proof, we compute the derivatives of  $M$  with respect to  $\alpha$  for arbitrary  $J$ . For  $j \geq i$ , the derivatives of  $\pi_j$  after  $\alpha_i$  equal

$$\begin{aligned}
\frac{d\pi_j}{d\alpha_i} &= -(\theta - \tau_i) \left( \exp \left( \sum_{s=1}^j \alpha_s (\theta - \tau_s) \right) \left( \sum_{k=i}^J \exp \left( \sum_{s=1}^k \alpha_s (\theta - \tau_s) \right) - z \right) \right) / z^2 \\
&= -(\theta - \tau_i) \pi_j \left( \sum_{k=i}^J \pi_k - 1 \right) = -(\theta - \tau_i) \left( \pi_j \left( \sum_{k=i}^J \pi_k \right) - \pi_j \right).
\end{aligned} \tag{34}$$

For  $j < i$ , the derivatives equal

$$\begin{aligned}
\frac{d\pi_j}{d\alpha_i} &= -(\theta - \tau_i) \left( \exp \left( \sum_{s=1}^j \alpha_s (\theta - \tau_s) \right) \left( \sum_{k=i}^J \exp \left( \sum_{s=1}^k \alpha_s (\theta - \tau_s) \right) \right) \right) / z^2 \\
&= -(\theta - \tau_i) \pi_j \left( \sum_{k=i}^J \pi_k \right).
\end{aligned} \tag{35}$$

Thus, we find

$$\sum_{j=1}^J \frac{d\pi_j}{d\alpha_i} = -(\theta - \tau_i) \left( \left( \sum_{j=1}^J \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J \pi_j \right), \tag{36}$$

and hence

$$\begin{aligned}
\frac{dM}{d\alpha_i} &= \sum_{j=1}^J \left( 2A_j \frac{dA_j}{d\alpha_i} \pi_j + A_j^2 \frac{d\pi_j}{d\alpha_i} \right) - 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{j=1}^J \left( \frac{dA_j}{d\alpha_i} \pi_j + A_j \frac{d\pi_j}{d\alpha_i} \right) \right) \\
&= -(\theta - \tau_i) \left( \left( \sum_{j=1}^J A_j^2 \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J A_j^2 \pi_j \right. \\
&\quad \left. - 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \sum_{j=i}^J A_j \pi_j \right) \right) \\
&\quad + \sum_{j=i}^J 2A_j \pi_j - 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{j=i}^J \pi_j \right).
\end{aligned} \tag{37}$$

Using Proposition 2(b), (37) evaluated at point  $\tau^*$  is equal to

$$\frac{dM}{d\alpha_i}(\tau^*, \alpha) = A_j \pi_j \pi_0 (2 - A_j (\theta - \tau_i^*) (\pi_j - \pi_0)), \tag{38}$$

which, for  $J = 1$  (dichotomous 2PL model;  $A_j = \alpha \in \mathbb{R}_+$ ), can be written as

$$\frac{dM}{d\alpha}(0, \alpha) = \alpha\pi_1\pi_0(2 - \alpha\theta(\pi_1 - \pi_0)). \quad (39)$$

Due to symmetry of  $\pi$  in the sense of Proposition 2(d) and since  $\Pi$  is symmetric, we have

$$- \int \theta\pi_1(\pi_1\pi_0)^x \Pi d\theta = \int \theta\pi_0(\pi_1\pi_0)^x \Pi d\theta \quad (40)$$

for any  $x \in \mathbb{Z}$ . For symmetric  $\Pi$ , the integral  $\int \pi_1\pi_0 \Pi d\theta$  is monotonically decreasing in  $\alpha$ , whereas

$$I_0(\alpha) := \int \theta\pi_1 \Pi d\theta \quad (41)$$

is monotonically increasing, which both is immediately evident from the graph of  $\pi_1$  and  $\pi_0$  changing with  $\alpha$ . For  $\psi_0$ , we have

$$\frac{d\psi_0}{d\alpha}(0, \alpha) = \int \frac{dM}{d\alpha}(0, \alpha) M(0, \alpha)^{-1} \Pi d\theta = \frac{2}{\alpha} \int (1 - \alpha\theta\pi_1) \Pi d\theta. \quad (42)$$

Since  $I_0(\alpha)$  is monotonically increasing in  $\alpha$ ,  $\alpha I_0(\alpha)$  is monotonically increasing and unbounded. Furthermore,  $\lim_{\alpha \rightarrow 0} \alpha I_0(\alpha) = 0$  and hence there exists a unique solution  $\alpha_0^*$  of

$$\alpha \int \theta\pi_1 \Pi d\theta = 1, \quad (43)$$

which (together with  $\tau^*$ ) constitutes the optimal one-point design for  $\psi_0$ . For  $\psi_{-1}$ , we have

$$\frac{d\psi_{-1}}{d\alpha}(0, \alpha) = \int \frac{dM}{d\alpha}(0, \alpha) M(0, \alpha)^{-2} \Pi d\theta = \frac{2}{\alpha^3} \int \frac{(1 - \alpha\theta\pi_1)}{\pi_1\pi_0} \Pi d\theta. \quad (44)$$

The integral

$$I_{-1}(\alpha) := \int \frac{(1 - \alpha\theta\pi_1)}{\pi_1\pi_0} \Pi d\theta \quad (45)$$

is monotonically decreasing in  $\alpha$  for symmetric  $\Pi$ , which again can be inferred from the graph of  $\pi_1$  and  $\pi_0$  changing with  $\alpha$ . Since  $\lim_{\alpha \rightarrow 0} I_{-1}(\alpha) = 4$  and  $\lim_{\alpha \rightarrow \infty} I_{-1}(\alpha) = -\infty$ , there exists a unique solution  $\alpha_{-1}^*$  of

$$\alpha \int \frac{\theta}{\pi_0} \Pi d\theta = \int \frac{1}{\pi_1\pi_0} \Pi d\theta, \quad (46)$$

which (together with  $\tau^*$ ) constitutes the optimal one-point design for  $\psi_{-1}$ .  $\square$

*Proof.* (Theorem 2) For a symmetric weight distribution, the first derivatives of  $\phi_0$  and  $\phi_{-1}$  are zero at point  $\tau^*$ , which can be seen directly from the first derivatives of  $\psi_0$  and  $\psi_{-1}$  in  $\tau^*$ . For  $i \geq n$ , we compute the elements of the Hessian matrix of  $M$  as

$$\begin{aligned}
\frac{d^2M}{d\tau_i d\tau_n} &= \alpha_i \alpha_n \left( \left( \left( \sum_{j=1}^J A_j^2 \pi_j \right) \left( \sum_{k=n}^J \pi_k \right) - \left( \sum_{j=n}^J A_j^2 \pi_j \right) \right) \left( \sum_{k=i}^J \pi_k \right) \right. \\
&+ \left( \sum_{j=1}^J A_j^2 \pi_j \right) \left( \left( \sum_{k=i}^J \pi_k \right) \left( \sum_{v=n}^J \pi_v \right) - \left( \sum_{k=i}^J \pi_k \right) \right) \\
&- \left( \sum_{j=i}^J A_j^2 \pi_j \right) \left( \sum_{k=n}^J \pi_k \right) - \left( \sum_{j=i}^J A_j^2 \pi_j \right) \\
&- 2 \left( \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{k=n}^J \pi_k \right) - \left( \sum_{j=n}^J A_j \pi_j \right) \right) \left( \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{k=i}^J \pi_k \right) - \left( \sum_{j=i}^J A_j \pi_j \right) \right) \\
&- 2 \left( \sum_{j=1}^J A_j \pi_j \right) \left( \left( \left( \sum_{j=1}^J A_j \pi_j \right) \left( \sum_{k=n}^J \pi_k \right) - \left( \sum_{j=n}^J A_j \pi_j \right) \right) \left( \sum_{k=i}^J \pi_k \right) \right. \\
&\quad \left. + \left( \sum_{j=1}^J A_j \pi_j \right) \left( \left( \sum_{k=i}^J \pi_k \right) \left( \sum_{v=n}^J \pi_v \right) - \left( \sum_{k=i}^J \pi_k \right) \right) \right) \\
&\quad \left. - \left( \sum_{j=i}^J A_j \pi_j \right) \left( \sum_{k=n}^J \pi_k \right) - \left( \sum_{j=i}^J A_j \pi_j \right) \right).
\end{aligned} \tag{47}$$

This holds for  $i < n$  as well due to symmetry of the Hessian matrix. Equation (47) can be slightly simplified, but remains too complicated to determine analytically under what conditions the Hessian matrix of  $\phi_v$  ( $v \in \{0, -1\}$ ) is positive definite for some Bayes optimal one-point design  $\tau_\Pi$  for a weight distribution  $\Pi$  of unspecified form. If, however,  $\Pi$  is symmetric we know from Lemma 1 that  $\tau^*$  is the Bayes optimal one-point design and (47) simplifies considerably to

$$\frac{d^2M}{d\tau_i d\tau_n}(\tau^*, \alpha) = \alpha_i \alpha_n \pi_j \pi_0 A_j^2 (1 - 6\pi_j \pi_0) = \alpha_i \alpha_n M(\tau^*, \alpha) (1 - 6\pi_j \pi_0), \tag{48}$$

and hence

$$\frac{d^2\phi_0}{d\tau_i d\tau_n}(\tau^*, \alpha) = \int M(\tau^*, \alpha)^{-1} \frac{d^2M}{d\tau_i d\tau_n}(\tau^*, \alpha) \Pi d\theta = \alpha_i \alpha_n \int (1 - 6\pi_j \pi_0) \Pi d\theta. \tag{49}$$

Within the design space of the threshold vector  $\tau$ , this result is of limited use since the  $\tau^*$  is at the border of the design region and the Hessian matrix is constant. However, considering that in  $\tau^*$  the GPCM reduces to the dichotomous 2PL model, the necessary condition for  $\tau^*$  to be Bayes optimal for  $v = 0$  can be directly inferred from (49) as

$$\int (1 - 6\pi_j\pi_0)\Pi d\theta \leq 0 \quad (50)$$

or, equivalently,

$$\int \pi_j\pi_0\Pi d\theta \geq \frac{1}{6}.$$

Similarly, when denoting  $C := (\int M(\boldsymbol{\tau}^*, \boldsymbol{\alpha})^{-1}\Pi d\theta)^{-1}$ , we obtain

$$\begin{aligned} \frac{d^2\phi_{-1}}{d\tau_i\tau_n}(\boldsymbol{\tau}^*, \boldsymbol{\alpha}) &= C \int M(\boldsymbol{\tau}^*, \boldsymbol{\alpha})^{-2} \frac{d^2M}{d\tau_i\tau_n}(\boldsymbol{\tau}^*, \boldsymbol{\alpha})\Pi d\theta \\ &= \alpha_i\alpha_n C \int (\pi_j\pi_0 J^2)^{-1} (1 - 6\pi_j\pi_0)\Pi d\theta \\ &= \alpha_i\alpha_n \left( C \int M(\boldsymbol{\tau}^*, \boldsymbol{\alpha})^{-1}\Pi d\theta - C \frac{6}{J^2} \int \Pi d\theta \right) \\ &= \alpha_i\alpha_n \left( 1 - C \frac{6}{J^2} \right). \end{aligned} \quad (51)$$

Thus, the necessary condition for  $\boldsymbol{\tau}^*$  to be Bayes optimal for  $\nu = -1$  becomes

$$6 \geq \int (\pi_j\pi_0)^{-1}\Pi d\theta. \quad (52)$$

□

*Proof.* (Theorem 3) This can be proved in the same way as Theorem 2 in Graßhoff et al. (2012), so we do not spell out the details here. □