Listening. Learning. Leading.®

Research Report
ETS RR–15-20

Location Indices for Ordinal Polytomous Items Based on Item Response Theory

Usama S. Ali

Hua-Hua Chang

Carolyn J. Anderson

December 2015

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Location Indices for Ordinal Polytomous Items Based on Item Response Theory

Usama S. Ali,[1] Hua-Hua Chang,[2] & Carolyn J. Anderson[2]

1 Educational Testing Service, Princeton, NJ

2 University of Illinois at Urbana-Champaign, Champaign, IL

Polytomous items are typically described by multiple category-related parameters; situations, however, arise in which a single index is needed to describe an item's location along a latent trait continuum. Situations in which a single index would be needed include item selection in computerized adaptive testing or test assembly. Therefore single location indices for ordinal polytomous items are proposed and studied. The proposed location indices (LIs) for polytomous items are mathematically derived based on the item category response functions (ICRFs) and item response function (IRF) for polytomous items. The ICRF approach resulted in three indices: LI_{mean}, $LI_{trimmed\ mean}$, and LI_{median}, and the IRF approach resulted in one proposed index, LI_{IRF}. An empirical example of real items is presented to help comprehension of the new location indices. Possible testing applications where the proposed item location indices are useful are discussed.

Keywords graded-response model; item response function; item response theory; location index; partial credit models; polytomous items

doi:10.1002/ets2.12065

Introduction

Items are the building blocks of psychological and educational tests, and the characteristics of the items determine the properties of the test. Item response theory (IRT) models specify the characteristics of items by parameters that are estimated from observed responses to items. These parameters act as descriptive statistics for the items. Researchers often use polytomous items for a variety of reasons, but mainly because these formats are more informative and reliable than dichotomously scored items. IRT models for polytomous items include the graded-response model (GRM; Samejima, 1969), partial credit model (PCM; Masters, 1982), generalized partial credit model (GPCM; Muraki, 1992), and nominal response model (Bock, 1972). In IRT models for dichotomous items, a single parameter conveys an item's difficulty; however, with a few exceptions, most of the previously mentioned polytomous models do not have a location parameter or location index associated with their standard definition. This presents a problem for using polytomous items in test assembly, in which typically a single index is required to determine what items to put on a test. Therefore this report discusses several variants of location indices that can be used with polytomous models. For a number of IRT Rasch models for polytomous items, item locations are well defined, such as the rating scale and successive interval models (see, e.g., Andrich, 1978, 1982; Rost, 1988).

Let $X_i = 0, 1, \ldots, m$ be the scores for ordinal, polytomous item i with corresponding probabilities $P_{i0}(\theta), \ldots, P_{im}(\theta)$ given a latent trait value θ. The probabilities, $P_{ix}(\theta)$, for different levels of θ are considered the item category response functions (ICRFs) where particular IRT models specify a (logistic) regression with a location or intercept parameter for response options as well as a slope for θ. Because the ICRFs have at least m parameters gauging difficulty and summarizing or describing the location of the item on the latent trait, defining an index that represents an item's location along the underlying continuum is not straightforward. As explained in later sections, the slope parameter for θ, a_i does not represent the discrimination of a polytomous item as it does in IRT models for dichotomous items. The slope parameter, a_i, combined with other category parameters defines the polytomous item discrimination (Embretson & Reise, 2000). The structure of polytomous IRT models leads to challenges in interpreting these parameters. For a dichotomous item, a single curve, an item characteristic curve or item response function (IRF), depicts the relationship between the latent variable

Corresponding author: U. S. Ali, E-mail: uali@ets.org

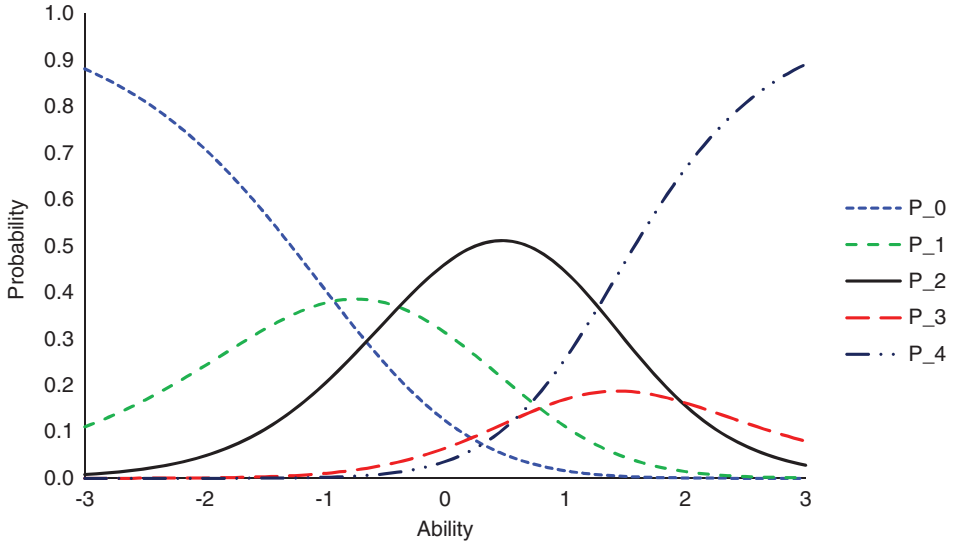

Figure 1 Item category response functions for a five-category item.

and the probability of correct response; however, for an item with $(m+1)$ response options, there exist $(m+1)$ ICRFs, one for each response category. Therefore it is essential to deal with $(m+1)$ curves to extract the information about item properties. As an example of the complexity, Figure 1 presents five ICRFs of an item with five response options.

For polytomous items, there is a distinction between an ICRF and an IRF. An ICRF gives the probability for a response option, whereas an IRF gives the expected value of X_i as a weight sum of the ICRFs; that is,

$$E\left(X_i\right) = P_{i1}\left(\theta\right) + 2P_{i2}\left(\theta\right) + \cdots + mP_{im}\left(\theta\right) = \sum_{x=1}^{m} xP_{ix}\left(\theta\right). \tag{1}$$

Chang and Mazzeo (1994) addressed the issue about the correspondence between an IRF and sets of ICRFs for a polytomous item. This correspondence is automatically satisfied for all dichotomous models for which the correct and incorrect response curves for a dichotomous item (i.e., two ICRFs) can be summarized by one curve (i.e., when $m=1$ in Equation 1, the IRF becomes equivalent to the correct response curve) that carries all information about item properties, as shown in Equation 2:

$$P_{i1}\left(\theta\right) = E\left(X_i\right). \tag{2}$$

According to Chang and Mazzeo, the item structure of a polytomous item is uniquely determined by its IRF for most commonly used models, and therefore the shape of the IRF contains all the information about the item, and no information will be lost by studying only this single curve. However, their study neglected quantifying the location of the IRF. More specifically, if an IRF uniquely determines the item structure, it is important to identify a location parameter for the IRF. The current report is a continuation on such an effort to propose a single index for polytomous items based on polytomous IRT models.

To summarize, the motivation and rationale to search for a central or an overall location parameter is twofold: (a) the complexity of multiple and different parameterizations for a polytomous item even for the same model and (b) the lack of a global item location parameter, which prevents the use of polytomous items in many testing applications, such as the usage of certain item selection methods in adaptive testing where providing such methods in a polytomous case is a challenge. The difference between the dichotomous and polytomous items in terms of parameters is the basis of the current report. New item indices for polytomous items are defined using the properties of an item's IRF and ICRFs. To provide a basis for the new measures, an overview is given of the different but most commonly used model parameterizations for polytomous items. These polytomous item response models are designed to analyze items with ordered response options.

The remainder of this report is structured as follows. In the next section, we briefly review the IRT models for which global location indices are developed (i.e., the GRM; Samejima, 1969) and the partial credit models (hereinafter

referring to both PCM, Masters, 1982, and GPCM, Muraki, 1992). After reviewing these models, global item location indices are proposed, and their characteristics are studied. An extended example using items from the National Assessment of Educational Progress is presented, followed by a discussion of the prospects for the use of the new indices.

Polytomous Item Response Models

The GRM (Samejima, 1969) and the partial credit models (Masters, 1982; Muraki, 1992) share a main characteristic: They have the same discrimination for each of the response options for an item. As noted in the following, the models differ in other aspects.

Graded Response Model

Samejima (1969) proposed a model for items that are characterized as ordinal response categories (e.g., Likert-scale items). The GRM expresses the cumulative probability of getting at least a score x:

$$P_i^*(X \geq x) = P_{ix}^* = \frac{\exp\left(a_i\left(\theta - b_{ix}\right)\right)}{1 + \exp\left(a_i\left(\theta - b_{ix}\right)\right)}, \tag{3}$$

where $x = 1, 2, \ldots, m$, a_i is the item slope parameter, and b_{ix} is a threshold parameter representing the point along the θ scale at which examinees have .50 probability of responding in or above category x, and therefore the probability of responding in a specific category score is $P_{ix} = P_{ix}^* - P_{i,x+1}^*$. Because the probability of a response for a specific category is the difference between the cumulative probabilities of two adjacent scores, Samejima's GRM is considered a difference model in Thissen and Steinberg's (1986) classification of IRT models. Note that it is assumed in the GRM that $P_{i0}^* = 1$ and $P_{i,m+1}^* = 0$ and that the b_{iv} s are ordered such as $b_{i1} < b_{i2} < \ldots < b_{im}$.

Partial Credit Models

Two versions of PCMs are considered here: Masters's (1982) PCM and Muraki's (1992) GPCM. Masters's PCM is based on a different conceptualization than the GRM and hence has a different parameterization. The PCM is suited to model sums of binary responses that are not supposed to be stochastically independent (Verhelst & Verstralen, 2008), and it is considered an extension of the dichotomous Rasch model to the polytomous case. The PCM belongs to the adjacent-category models in Mellenbergh's (1995) classification of IRT models and to the divide-by-total models in Thissen and Steinberg's (1986) classification. As a model in the Rasch family, the PCM considers items to be equally discriminating (i.e., they all have the same slope, $a_i = a$). Allowing items to be differentially discriminating yields the GPCM (Muraki, 1992, 1993).

The GPCM for an examinee with ability θ states that the probability of getting a score x of item i denoted by $P(\theta)$ is

$$P_{ix}(\theta) = \frac{\exp \sum_{v=1}^{x} Da_i\left(\theta - b_{iv}\right)}{1 + \sum_{c=1}^{m} \exp \sum_{v=1}^{c} Da_i\left(\theta - b_{iv}\right)}, \tag{4}$$

where D is a scaling constant that puts the trait scale in the same metric as the normal ogive model ($D = 1.7$) or stays on the metric of the logistic model ($D = 1$) that will be used after this position, a_i is a slope parameter for item i, and b_{iv} are m threshold parameters. Special cases of the GPCM include Birnbaum's (1968) two-parameter logistic model when $m = 1$, the PCM when $a_i = a$, and the Rasch model for dichotomous items when $m = 1$ and $a_i = a$. It should be noted that in the PCMs, the thresholds need not to be ordered. Note that in the parameterization of the GPCM implemented in PARSCALE (Muraki & Bock, 2003), there is a location b_i and threshold distances $d_{iv} = b_i - b_{iv}$ that add up to 0. For example, for a three-category item, there is one location parameter and two threshold distances, where $d_{i1} = -d_{i2}$.

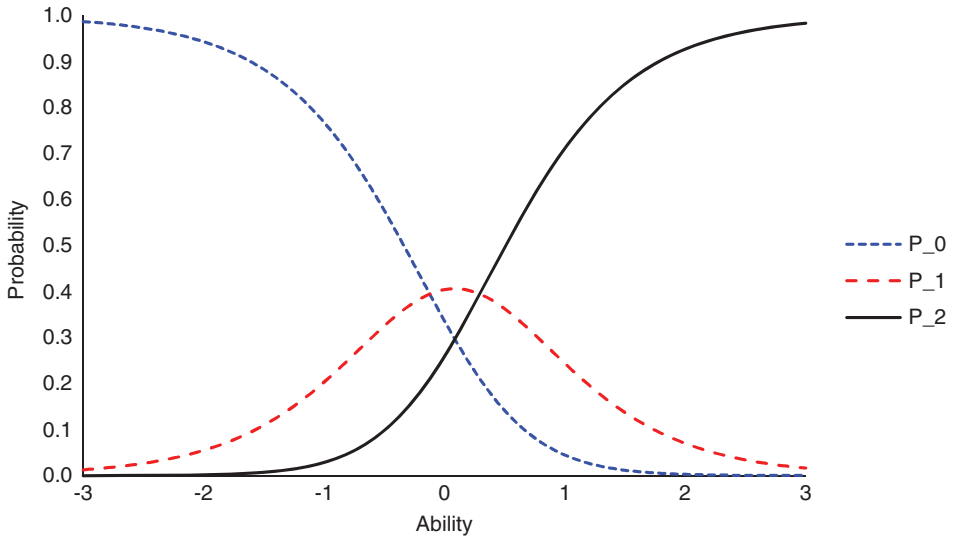

Figure 2 Item category response functions for a three-category item.

Proposed Global Item Location Indices

Two general approaches are used to develop item location indices for polytomous items. The first approach is to study the category response functions, and it will consider only the PCMs but not the GRM. The second one focuses on the IRF for all models. Basically, the proposed indices are based on the ICRFs and IRF of a polytomous item using the preceding models (Ali, 2011).

Indices Based on the Item Category Response Function

Using a partial credit model (i.e., PCM or GPCM), we developed indices for a polytomous item with three categories and subsequently generalized the indices to items with more response options. Consider a three-category item that follows the partial credit models (i.e., PCM and GPCM). The parameters for item i are a_i, b_{i1}, and b_{i2}. There are three ICRFs for the three possible scores 0, 1, or 2 on the item; that is, $P_{i0}(\theta)$, $P_{i1}(\theta)$, and $P_{i2}(\theta)$. Examples of three such ICRFs for a three-response option item based on a partial credit model are shown in Figure 2. Note that the ICRFs intersect with each other at some points; the zero- and perfect-score ICRFs intersect in a point corresponding to the peak of the partial credit ICRF. Thus the middle score ICRF plays a key role in the development of an item's location index. It can be shown mathematically that this will always occur. Using the probability of attaining the partial credit (middle) score, $P_{i1}(\theta)$, we locate the peak of this ICRF. From Equation 4,

$$P_{i1}(\theta) = \frac{\exp\left[a_i\left(\theta - b_{i1}\right)\right]}{1 + \exp\left[a_i\left(\theta - b_{i1}\right)\right] + \exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right]}, \tag{5}$$

and the first derivative of Equation 5 with respect to θ is

$$\begin{aligned}\frac{\partial P_{i1}(\theta)}{\partial \theta} &= a_i P_{i1}(\theta) - a_i P_{i1}(\theta)\sum_{c=1}^{2} cP_{ic}(\theta) \\ &= a_i P_{i1}(\theta)\left[1 - \sum_{c=1}^{2} cP_{ic}(\theta)\right].\end{aligned} \tag{6}$$

Setting Equation 6 equal to zero, we find that the maximum of Equation 6 occurs when any of the following three conditions hold: $a_i = 0$, $P_{i1}(\theta) = 0$, or $\sum_{c=1}^{2} cP_{ic}(\theta) = 1$ (or $E(X_i) = 1$). The first condition, $a_i = 0$, indicates that the item has no discrimination power, and hence it would not be in an operational item pool. The second condition, $P_{i1}(\theta) = 0$, is not achievable. All response options have nonzero probability. The third condition, $\sum_{c=1}^{2} cP_{ic}(\theta) = 1$ or $E(X_i) = 1$, can be

attained as seen by noting that

$$\begin{aligned}
\sum_{c=1}^{2} cP_{ic}(\theta) &= 1 \\
P_{i1}(\theta) + 2P_{i2}(\theta) &= 1 \\
\exp\left[a_i\left(\theta - b_{i1}\right)\right] + 2\exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right] &= 1 + \exp\left[a_i\left(\theta - b_{i1}\right)\right] + \exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right] \\
\exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right] &= 1 \\
a_i\left(2\theta - b_{i1} - b_{i2}\right) &= 0 \\
\theta &= \frac{1}{2}\left(b_{i1} + b_{i2}\right). \qquad (7)
\end{aligned}$$

With regard to the point on the ability continuum corresponding to the intersection between these two ICRFs of scores 0 and 2, it satisfies the following condition:

$$\begin{aligned}
P_{i0}(\theta) &= P_{i2}(\theta) \\
\frac{1}{1 + \exp\left[a_i\left(\theta - b_{i1}\right)\right] + \exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right]} &= \frac{\exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right]}{1 + \exp\left[a_i\left(\theta - b_{i1}\right)\right] + \exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right]}. \qquad (8)
\end{aligned}$$

The equivalence in Equation 8 implies that

$$\exp\left[a_i\left(2\theta - b_{i1} - b_{i2}\right)\right] = 1, \qquad (9)$$

which is the same conclusion as given in Equation 7; that is, $\theta = \frac{1}{2}\left(b_{i1} + b_{i2}\right)$.

Because we verified that the two ICRFs for the lowest and highest scores on the item intersect, we note that this corresponds to the same point on the ability scale as the maximum of the ICRF of the partial credit score (see Figure 2 for an example of a three-category item).

For the more general case of a polytomous item with $m+1$ categories, where every two ICRFs of scores x and $m-x$ intersect in a point (i.e., $P_{i0}(\theta) = P_{im}(\theta)$), we have the following conditions: $(P_{i1}(\theta) = P_{i,m-1}(\theta), \ldots, P_{ix}(\theta) = P_{i,m-x}(\theta), \ldots, P_{i,\frac{m+1}{2}}(\theta) = P_{i,\frac{m+3}{2}}(\theta))$. See Figure 1 for an example of a five-category item. Following the same logic used to obtain the result for the three-category item, we start with two curves at their point of intersection; that is,

$$P_{ix}(\theta) = P_{i,m-x}(\theta).$$

Substituting the model for each of the functions and simplifying yields

$$\begin{aligned}
x\theta - \sum_{c=1}^{x} b_{ic} &= (m-x)\theta - \sum_{c=1}^{m-x} b_{ic}, \\
[(m-x) - x]\theta &= \sum_{c=1}^{m-x} b_{ic} - \sum_{c=1}^{x} b_{ic}, \\
(m-2x)\theta &= \sum_{c=x+1}^{m-x} b_{ic}. \qquad (10)
\end{aligned}$$

This last relation suggests that a reasonable location index is based on

$$\theta = \frac{1}{m-2x}\sum_{c=x+1}^{m-x} b_{ic}, \qquad x = 0, 1, \ldots, \frac{m+1}{2}. \qquad (11)$$

In other words, at the two middle ICRFs, $\theta = b_{i,(m-1/2)}$. When m is an even integer, as represented by the five-category item example in Figure 1, such that there is one middle ICRF representing the score of $m/2$, we need a point that corresponds to the maximum of this ICRF.

To conclude, Table 1 summarizes the relationship of category characteristic curves of a polytomous item with ordered response options scored 0 to m and the formula of the corresponding intersection points on the ability scale with reference

Table 1 Studied Item Category Response Functions and Corresponding Intersection Points

ICRFs	Intersection point	Notes
C_0, C_1	b_1	Model definition
C_1, C_2	b_2	Model definition
C_{x-1}, C_x	b_x	Model definition
C_0, C_2	$0.5(b_1 + b_2)$	The same as the peak of C_1 for a three-category item
C_x, C_{m-x}	$(m-2x)^{-1}\sum_{c=x+1}^{m-x} b_{ic}$	General form of intersecting point of x and $m-x$ score curves
C_x, C_y	$\left(m-x-y\right)^{-1}\sum_{c=x+1}^{m-y} b_{ic}$	More general form of intersecting point of any two curves
C_0, C_m	$m^{-1}\sum_{c=1}^{m} b_{ic}$	General form for the intersecting point of 0 and m score curves

Note. C_v = item characteristic curve for score v.

to the definition of such scale values. This overall summary of the relations among ICRFs suggests the following proposed location indices. On the basis of the preceding mathematical derivation, we propose alternative forms of a location index (LI) for a polytomous item.

Proposal 1

The first form of an LI is the average item category difficulties, which takes all ICRFs into account ($\mathrm{LI}_{\mathrm{mean}}$) by substituting $x = 0$ into Equation 11:

$$\mathrm{LI}_{\mathrm{mean}} = \frac{1}{m}\sum_{c=1}^{m} b_{ic}. \tag{12}$$

Note that this proposed index is very similar to item location of Andrich's rating scale model and it is also the location parameter b generated by PARSCALE (Muraki & Bock, 2003).

Proposal 2

The second form of LI is the median of item category difficulties ($\mathrm{LI}_{\mathrm{median}}$), which is a possible choice in statistics, as follows:

$$\mathrm{LI}_{\mathrm{median}} = \mathrm{Median}\left(b_{ix}s\right) = \begin{cases} b_{ix}^{(k)}, & \text{if } m \text{ is even,} \\ 0.5\left(b_{ix}^{(k)} + b_{ix^*}^{(k+1)}\right), & \text{if } m \text{ is odd,} \end{cases} \tag{13}$$

where $b_{ix}^{(k)}$ is the threshold parameter that has the kth rank among the thresholds of the ith item and has score x, $b_{ix^*}^{(k+1)}$ is the threshold parameter that has the $(k+1)$th, and $b_{ix}^{(k)} \leq b_{ix^*}^{(k+1)}$.

Proposal 3

The third form of an LI is the truncated (trimmed or Windsor) mean; that is, the average of item category difficulties that takes all ICRFs into account, except the zero- and perfect-score curves if there are no reversals ($\mathrm{LI}_{\mathrm{trimmed\ mean}}$), by substituting $x = 1$ into Equation 11:

$$\mathrm{LI}_{\mathrm{trimmed\ mean}} = \frac{1}{m-2}\sum_{c=2}^{m-1} b_{ic}. \tag{14}$$

In the case of reversal, the threshold parameters should be rank ordered in a similar way as in the previous index.

Index Based on the Item Response Function

The ease of calculating a polytomous IRF follows from the fact that an IRF can be thought of as describing the rate of change of expected value of an item response as a function of the change in θ relative to an item's location b_i (Nering &

Ostini, 2006). More succinctly, this can be thought as a regression of the item score onto the trait ability (Chang & Mazzeo, 1994; Lord, 1980).

The previous three proposals of LI are based on the ICRFs; hence they are considered as local indices by the nature of information gained from curves of specific score categories. Conversely, this is not the case in polytomous models. Chang and Mazzeo (1994) showed that the IRF for a polytomously scored item is defined as a weighted sum of the ICRFs (the probability of getting a particular score for a randomly sampled examinee of ability); that is, it is defined by the expected value or mean of the scores.

The IRF, as defined in Equation 1, ranges from 0 to m (i.e., the maximum possible score category of an item). They established the correspondence between an IRF and a unique set of ICRFs for two of the most commonly used polytomous IRT models (GRM and the partial credit models). Specifically, Chang and Mazzeo provided a proof for these models, as follows:

> If two items have the same IRF, then they must have the same number of categories; moreover, they must consist of the same ICRFs.

The condition on which the proof depends is that the discrimination parameter for each item does not depend on the category (i.e., for a given item, the a parameter is the same for each category or response option). The GRM, PCM, and GPCM all satisfy this condition, but the nominal response model does not satisfy it.

Along the same lines, Akkermans and Muraki (1997) introduced an IRF defined as a normalized expected score (i.e., the weighted sum of ICRFs divided by the number of item categories) that ranges from 0 to 1. Akkermans and Muraki's IRF differs in terms of the range from that introduced by Chang and Mazzeo (1994). Akkermans and Muraki (1997) introduced the gradient (i.e., first derivative) of IRF as an item discrimination function, $G(\theta)$,

$$G_i(\theta) = \frac{\partial \overline{T}_i(\theta)}{\partial \theta} = a_i^2 \left[\sum_{x=1}^{m} x^2 P_{ix}(\theta) - \left(\sum_{x=1}^{m} x P_{ix}(\theta) \right)^2 \right] = \frac{I_i(\theta)}{a_i}, \tag{15}$$

where $\overline{T}_i(\theta)$ is the IRF, where $T_i(\theta)$ is called the scoring function (Andrich, 1978), and $I_i(\theta)$ is the item information (for the specific formulas for different models, see Dodd, de Ayala, & Koch, 1995; Muraki, 1993; Nering & Ostini, 2010).

The polytomous IRF has various merits. First, the IRF carries the full information of the item and encompasses the partial amount of information included in ICRFs. Second, it is valid to apply the expected score to the most commonly used ordinal response models (i.e., GRM, PCM, and GPCM). Third, the IRF is well connected to Fisher information (see Equation 15). For the preceding properties of the IRF or expected score of a polytomous item, it is worthwhile to use it to propose a central location parameter.

Proposal 4

The fourth form of LI is derived from the polytomous IRF. Binary IRT models such as one-, two-, or three-parameter logistic models have an important feature in that the conditional mean of the item score (i.e., expectation) is the same as the probability of answering the item correctly. Note that the dichotomous IRF uses the value of .05 (if there is no guessing) as a threshold to determine the item location where the highest score is 1. Using the same analogy, the index of a polytomous IRF corresponds to an expected score of $0.5m$, where m is the highest possible score for an $(m+1)$-response category item. Because this value has a global nature in that it considers the IRF, we call it $\mathrm{LI}_{\mathrm{IRF}}$:

$$\mathrm{LI}_{\mathrm{IRF}} = \theta : E\left[X_i\right] = \frac{m}{2}. \tag{16}$$

For example, the θ point that corresponds to the $0.5m$ under the partial credit models can be obtained through the following equation, in which the closed-form solution is complicated to produce

$$\sum_{x=1}^{m} \left[(2x - m) \exp\left(a_i \left(x\theta - \sum_{c=1}^{x} b_{ic} \right) \right) \right] = m. \tag{17}$$

An iterative algorithm is used to obtain the $\mathrm{LI}_{\mathrm{IRF}}$ for each polytomous item. Here are the details of the Newton – Raphson method to obtain the approximate value of $\mathrm{LI}_{\mathrm{IRF}}$ for both the partial credit models and the GRM.

The Newton – Raphson method is a numerical method to solve nonlinear equations of the form of $f(x) = 0$. The approximate solution to the equation is

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}, \tag{18}$$

where x_{t+1} is the updated approximation based on the previous estimate, x_t, and $f'(x_t)$ is the first derivative of $f(x_t)$ with respect to x.

In the following section, we introduce the approximation of $\mathrm{LI}_{\mathrm{IRF}}$ using the three different polytomous IRT models: the partial credit models (Masters, 1982; Muraki, 1992) and the GRM (Samejima, 1969).

$\mathrm{LI}_{\mathrm{IRF}}$ of the Partial Credit Models' Items

In the current case, consider a polytomous item with $m+1$ response categories ranging from 0 to m. The formula for obtaining a category x on item i using a general form to the partial credit models is given by Equation 4. The expected score given a specific ability value θ is given by Equation 1. Assuming that $m/2$ is our critical point to get the corresponding θ value that satisfies such a criterion, we have the following:

$$\sum_{x=1}^{m} xP_{ix}(\theta) = \frac{m}{2}. \tag{19}$$

Therefore the function that needs to be solved is

$$f(\theta) = \sum_{x=1}^{m} xP_{ix}(\theta) - \frac{m}{2} = 0. \tag{20}$$

Given that

$$\frac{\partial P_{ix}(\theta)}{\partial \theta} = a_i P_{ix}(\theta) \left[x - \sum_{c=1}^{m} cP_{ic}(\theta) \right], \tag{21}$$

the first derivative of $f(\theta)$ with respect to θ, $f'(\theta)$, is given by

$$f'(\theta) = \sum_{x=1}^{m} x a_i P_{ix}(\theta) \left[x - \sum_{c=1}^{m} cP_{ic}(\theta) \right]. \tag{22}$$

The approximate value of $\mathrm{LI}_{\mathrm{IRF}}$ using the partial credit models is

$$\begin{aligned} \theta_{t+1} &= \theta_t - \frac{f(\theta_t)}{f'(\theta_t)} \\ &= \theta_t - \frac{\left[\sum_{x=1}^{m} xP_{ix}(\theta_t) \right] - \frac{m}{2}}{\sum_{x=1}^{m} x a_i P_{ix}(\theta_t) \left[x - \sum_{c=1}^{m} cP_{ic}(\theta_t) \right]}. \end{aligned} \tag{23}$$

$\mathrm{LI}_{\mathrm{IRF}}$ of the Graded Response Model's Items

The formula for obtaining a category x on item i using a general form to the graded response model is given by

$$P_{ix}(\theta) = P^*_{ix}(\theta) - P^*_{i,x+1}(\theta), \tag{24}$$

where $P^*_{ix}(\theta)$ is given by the formula of a two-parameter logistic model, as follows:

$$P^*_{ix}(\theta) = \frac{1}{1 + \exp\left[-a_i(\theta - b_{ix})\right]}. \tag{25}$$

Therefore the function that needs to be solved is

$$f(\theta) = \sum_{x=1}^{m} x \left(P_{ix}^{*}(\theta) - P_{i,x+1}^{*}(\theta) \right) - \frac{m}{2} = 0. \tag{26}$$

Given that

$$\sum_{x=1}^{m} x \left(P_{ix}^{*}(\theta) - P_{i,x+1}^{*}(\theta) \right) = \sum_{x=1}^{m} P_{ix}^{*}(\theta) \tag{27}$$

and

$$\frac{\partial P_{ix}^{*}(\theta)}{\partial \theta} = a_i P_{ix}^{*}(\theta) \left[1 - P_{ix}^{*}(\theta) \right], \tag{28}$$

the first derivative of $f(\theta)$ with respect to θ, $f'(\theta)$, is given by

$$f'(\theta) = \sum_{x=1}^{m} a_i \left[P_{ix}^{*}(\theta) \left(1 - P_{ix}^{*}(\theta) \right) \right]. \tag{29}$$

The approximate value of $\mathrm{LI}_{\mathrm{IRF}}$ using the GRM is

$$\begin{aligned} \theta_{t+1} &= \theta_t - \frac{f\left(\theta_t\right)}{f'\left(\theta_t\right)} \\ &= \theta_t - \frac{\left[\sum_{x=1}^{m} P_{ix}^{*}\left(\theta_t\right) \right] - \frac{m}{2}}{\sum_{x=1}^{m} a_i \left[P_{ix}^{*}\left(\theta_t\right) \left(1 - P_{ix}^{*}\left(\theta_t\right) \right) \right]}. \end{aligned} \tag{30}$$

For a given polytomous item with three response categories, there is a correspondence between the $\mathrm{LI}_{\mathrm{IRF}}$ and the ICRFs-based LIs (see Equation 7 and Figures 2 and 3). For items with more than three response categories, the values of these indices are different (see Figures 1 and 4).

An Extended Example

The following empirical example is an illustration of computing the LIs for a polytomous item. Table 2 provides GPCM parameters for five four-category items and three six-category items from the National Assessment of Educational Progress. The corresponding LI for each item is calculated using the formulas presented in the previous sections.

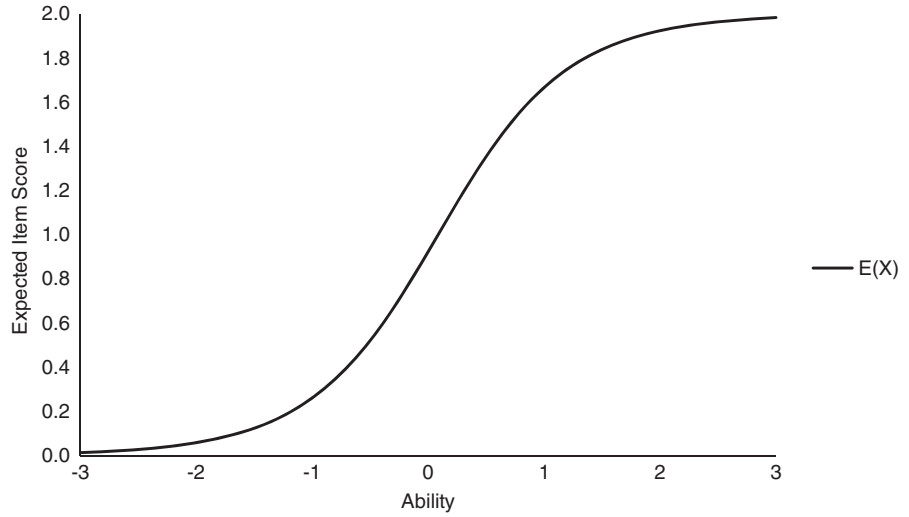

Figure 3 Item response function for a three-category item.

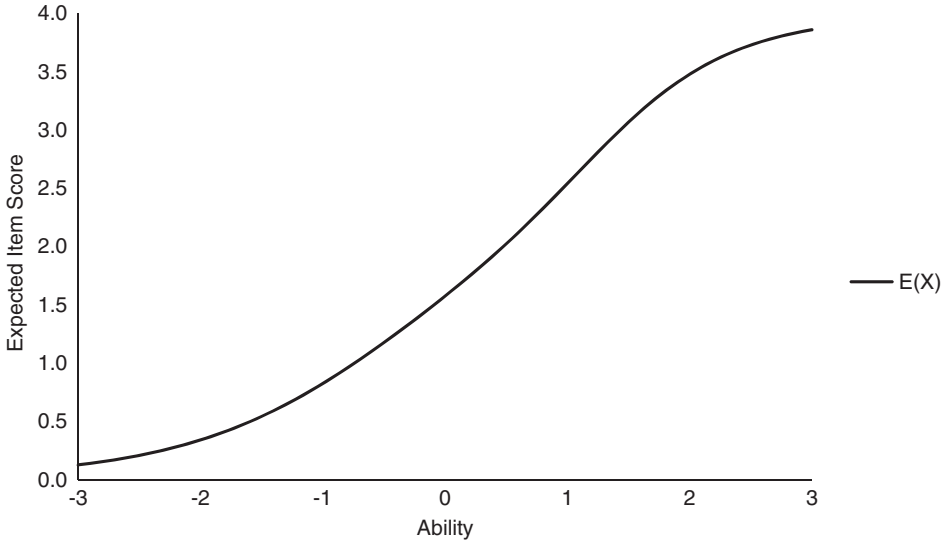

Figure 4 Item response function for a five-category item.

Table 2 Item Parameters and the Corresponding Location Indices

	GPCM parameters						Location indices			
Item *i*	a_i	b_{i1}	b_{i2}	b_{i3}	b_{i4}	b_{i5}	LI_{mean}	$LI_{trimmed\ mean}$	LI_{median}	LI_{IRF}
1	0.600	−0.490	0.810	1.770			0.700	0.810	0.810	0.720
2	0.680	−0.300	0.900	1.830			0.810	0.900	0.900	0.830
3	0.620	−0.390	−0.590	0.230			−0.250	−0.390	−0.390	−0.290
4	1.170	0.360	1.210	1.640			1.070	1.210	1.210	1.100
5	0.640	−0.170	−0.200	0.610			0.210	−0.170	−0.170	0.050
6	0.164	13.407	−7.203	−1.454	2.022	4.858	2.326	1.809	−2.022	1.468
7	0.465	−4.274	−0.21	0.611	−0.343	0.478	−1.054	−0.025	−0.210	−0.305
8	0.296	−0.287	−1.432	0.905	−2.970	−1.414	−0.946	−1.044	−1.414	−1.029

As an example, the parameters for Item 1 are $a_1 = 0.60$, $b_{11} = -0.49$, $b_{12} = 0.81$, and $b_{13} = 1.77$. Therefore LI_{mean} $LI_{mean} = \sum(b_{1x}/3) = (-0.49 + 0.81 + 1.77)/3 = 0.70$, and $LI_{trimmed\ mean} = LI_{median} = b_{12} = 0.81$. Regarding the LI_{IRF} for Item 1, it is computed using Equation 23. Suppose we choose $\theta_0 = 0$ as an estimate for the LI_{IRF} so we can update this estimate using Equation 23, where $m = 3$. Hence after two iterations, $LI_{IRF} = 0.72$.

Because the four proposed LIs are identical for the case of three-category items, such items are not reported in Table 2. From the table, it is obvious the $LI_{trimmed\ mean}$ and LI_{median} are equal for the four-category items (e.g., Items 1–5). For items with more than five response options, these two LIs are different (e.g., Items 6–8).

From the nature of the different proposals for the LI, we can infer some results. One result is that the LI_{mean} does not reflect the order of the thresholds, or b_{ix}; its value stays the same even though these parameters differ in their rank order (i.e., reversals). Alternating the values attached to these categories will not affect any of the ICRF-based indices. This characteristic holds for both the $LI_{trimmed\ mean}$ and LI_{median}. With regard to the LI_{IRF}, it has a sound foundation gained from the IRF, and therefore it is more appropriate to represent the difficulty of a polytomous item as a whole.

Discussion

Global indices that reflect the location of a polytomous item with ordered-response options were proposed. The methodology depending on a polytomous item's multiple response curves (i.e., ICRFs) or a single expected curve (i.e., IRF) is also valid for dichotomous items. In other words, the difficulty parameter for a dichotomous item can be obtained using either approach (i.e., studying the two response curves or the correct response curve). To illustrate using a two-parameter

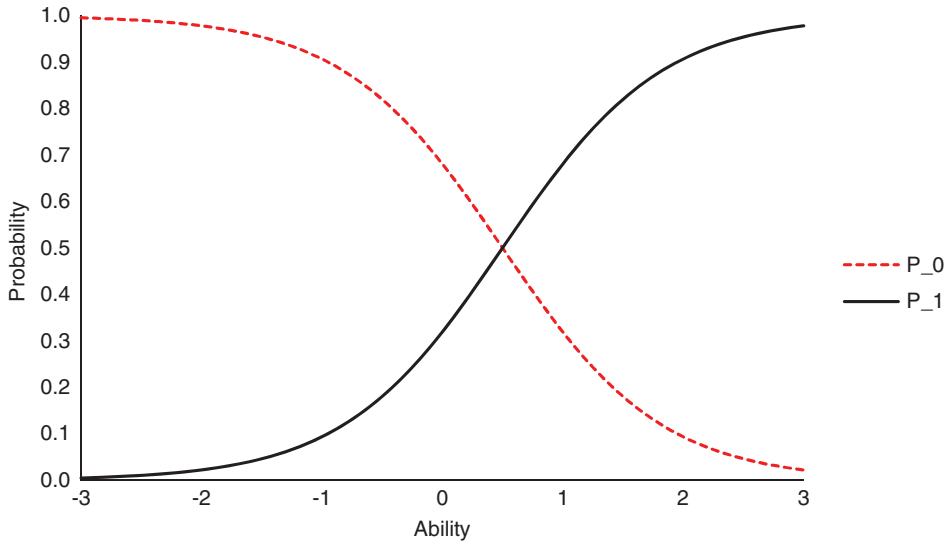

Figure 5 Item category response functions for a two-category item.

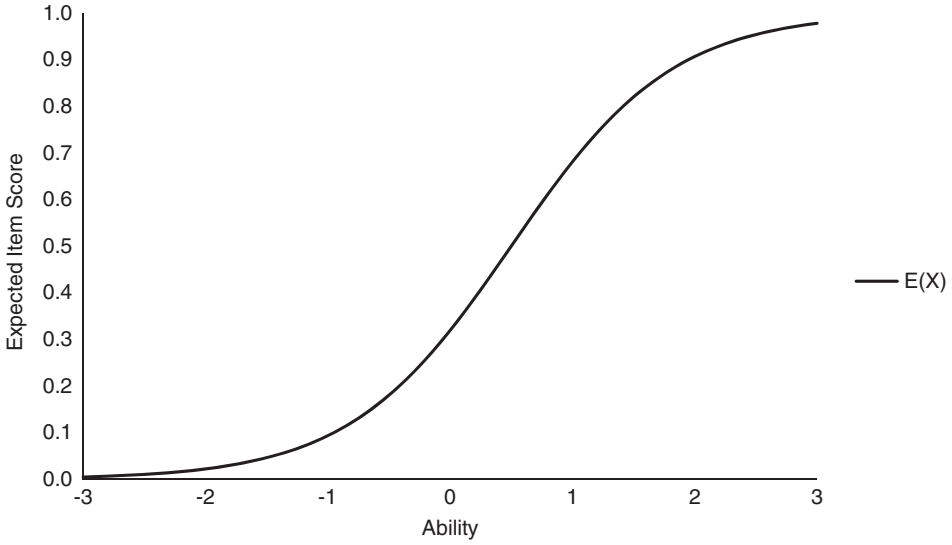

Figure 6 Expected score for a two-category item.

logistic model (Lord & Novick, 1968) to model responses of a dichotomous item, we can have the intersection point of the two characteristic curves of such an item (i.e., the curves that correspond to correct and incorrect answers) at the difficulty parameter, b_i. Figure 5 shows an example of the curves of correct and incorrect responses of a dichotomous item intersecting in a point corresponding to $\theta = 1.0$ (the item difficulty parameter). This figure is based on the two curves representing the ICRFs, the first approach used in the polytomous case. The three forms of an LI (i.e., LI_{mean}, $LI_{trimmed\ mean}$, and LI_{mean}) are simplified to produce one form for dichotomous items because they have only two response options (correct – incorrect or perfect – zero scores).

While based on the point of view of expected item score or IRF, the point on the ability scale corresponding to an expected score of .05 for dichotomous scoring represents an index of item difficulty. Figure 6 shows the correct-response curve of the same item that has a difficulty parameter of $b_i = 1.0$. This curve also represents the expected score conditional on ability level (i.e., the IRF), and it is obvious that the ability level corresponding to an expectation value of .50 is $\theta = 1.0$. This provides the basis for the second approach.

LIs of polytomous items have several potential uses. One potential use is in test assembly. An index representing an item's location along the latent variable replaces the multiple category parameters. By such an index, the characteristics of a polytomous item can also be represented by two main item parameters (i.e., a and the LI), in addition to the category-related parameters. Also, the number of categories is important; the area under the item information function summarizes the amount of information an item can give. For any GPCM item, for instance, this area is ma (Huynh, 1994, 1996). Hence, when tests are created using polytomous items, these two parameters per item would be useful for selecting items to satisfy target test specifications. Many testing programs use mixed-format tests in which both selected-response items and constructed-response items are included. One main target in test specifications of a test is the mean difficulty. In many cases, test developers only consider the section of selected-response items and ignore the other section of constructed-response items for different reasons such as they are few relative to the test length. In such cases, the LI of these ignored items can help test developers achieve the target test difficulty. Test assembly and construction of parallel forms can be easily done using the results of the current report in conjunction with Chang and Mazzeo's (1994) results. We have been able to quantify the IRF of a polytomous item into a single value that Chang and Mazzeo proved that, for any two items following any of the studied models, can be treated as equivalent, provided that their IRFs coincide.

A second potential area for using the LI is computerized adaptive testing (CAT). Based on the literature of testing with polytomous items, in particular adaptive testing, some item selection methods are natural extensions of those used with dichotomous items, including information indices. The information-based item selection approach may consider the item as a whole or at the score category level. Dodd et al. (1995) commented that only the information-based item selection algorithms have been investigated for the GRM and PCM because there is no single index or summary of the multiple location (or scale value) parameters in these models. In the context of adaptive testing, the item selection approach based on matching the difficulty can be presented such that an individual's estimated ability level is matched to a polytomous item's LI. In particular, four proposed item selection methods in polytomous adaptive testing are built based on the alternative forms of the polytomous-item LI. The choice of the next item to be administered is based on each form of the proposed index that matches the current ability estimate. For example, considering the $\mathrm{LI}_{\mathrm{IRF}}$ computed for an item under a polytomous response model, the next item for administration is chosen based on matching $\mathrm{LI}_{\mathrm{IRF}}$ to the current estimate of an examinee's ability.

Lima Passos, Berger, and Tan (2008) presented some findings regarding the item's $(b_{i1} + b_{i2})/2$. This index, based on results reported in this report, corresponds to the mean of item category thresholds, $\mathrm{LI}_{\mathrm{mean}}$, where all LIs, such as $\mathrm{LI}_{\mathrm{IRF}}$ in the case of three-category items, are equal, as shown in Equation 7. Lima Passos et al. found that the smaller the difference given by $[(b_{i1} + b_{i2})/2] - \theta$, the better (i.e., the more accurate) the tailoring between a selected item i and the underlying trait θ is. This is the core idea of the matching LI procedure in polytomous adaptive testing and one of the main applications of polytomous item LIs. Other applications are possible, such as extending the idea of a single location index to item sets or testlets by proposing an overall testlet LI. How well these indices work for CAT and for test assembly remains a question for further study.

References

Akkermans, W., & Muraki, E. (1997). Item information and discrimination functions for trinary PCM items. *Psychometrika*, *62*, 569 – 578.

Ali, U. S. (2011). *Item selection methods in polytomous computerized adaptive testing* (Unpublished doctoral dissertation). University of Illinois, Urbana, IL.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561 – 573.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47*, 105 – 113.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397 – 479). Reading, MA: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *27*, 29 – 51.

Chang, H.-H., & Mazzeo, J. (1994). The unique correspondence of the item response function and item category response functions in polytomously scored item response models. *Psychometrika*, *59*, 391 – 404.

Dodd, B. G., de Ayala, R. J., & Koch, W. R. (1995). Computerized adaptive testing with polytomous items. *Applied Psychological Measurement*, *19*, 5 – 22.

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

Huynh, H. (1994). On equivalence between a partial credit item and a set of independent Rasch binary items. *Psychometrika*, *59*, 111 – 119.

Huynh, H. (1996). Decomposition of a Rasch partial credit item into independent binary and indecomposable trinary items. *Psychometrika*, *61*, 31 – 39.

Lima Passos, V., Berger, M. P. F., & Tan, F. E. (2008). The D-optimality item selection criterion in the early stage of CAT: A study with the graded response model. *Journal of Educational and Behavioral Statistics*, *33*, 88 – 110.

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149 – 174.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, *19*, 91 – 100.

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

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, *17*, 351 – 363.

Muraki, E., & Bock, R. D. (2003). PARSCALE 4 for Windows: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks [Computer software]. Lincolnwood, IL: Scientific Software International.

Nering, M. L., & Ostini, R. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage.

Nering, M. L., & Ostini, R. (2010). New perspectives and applications. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 3 – 20). New York, NY: Routledge.

Rost, J. (1988). Measuring attitudes with a threshold model drawing on a traditional scaling concept. *Applied Psychological Measurement*, *12*, 397 – 409.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometrika Monograph No. 17). Richmond, VA: Psychometric Society.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*, 567 – 577.

Verhelst, N. D., & Verstralen, H. H. F. M. (2008). Some considerations on the partial credit model. *Psicológica*, *29*, 229 – 254.

Suggested citation:

Ali, U. S., Chang, H.-H., & Anderson, C. J. (2015). *Location indices for ordinal polytomous items based on item response theory* (Research Report No. RR-15-20). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12065

Action Editor: Matthias von Davier

Reviewers: Shelby Haberman, John R. Donoghue, and Peter van Rijn