Information Functions of the Generalized Partial Credit Model

Eiji Muraki

Educational Testing Service

The concept of information functions developed for dichotomous item response models is adapted for the partial credit model. The information function is explained in terms of the model parameters and scoring functions. The relationship between the item information function and the item response function also is discussed. The information function then is used to investigate the effect of collapsing and recoding categories of polytomously-scored items of the National Assessment of Educational Progress (NAEP). The NAEP writing items were calibrated and the item and test information is used to discuss desirable properties of polytomous items.   *Index terms: information function, item response model, National Assessment of Educational Progress (NAEP), partial credit model, polytomous item response model.*

Tests are constructed to measure examinee trait levels. Because the information function (from item response theory) provides valuable information about the precision of measurement at a specified trait level, the information function has been widely used for various applications that are central to the field of measurement (Hambleton & Swaminathan, 1985). The applications of the information function for dichotomously scored test items have been studied extensively, but applications of the information function for polytomous item response models have not been fully explored. The partial credit model (PCM) is frequently used to analyze polytomous cognitive and affective response data (Masters, 1982; Muraki, 1992; Wright & Masters, 1982). In this paper, the advantages of the PCM are demonstrated and compared to the dichotomous item response model using information functions. Some cautions about the uncritical application of the PCM also are provided.

The Generalized PCM

The generalized PCM (GPCM; Muraki, 1992) is formulated on the assumption that the probability of selecting the kth category over the $k - 1$st category in a multicategory item is governed by the logistic dichotomous response model; that is,

$$P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[Da_j(\theta - b_{jk})]}{1 + \exp[Da_j(\theta - b_{jk})]} \, , \tag{1}$$

where $k = 2, 3, \ldots, m_j$, and m_j is the number of response categories. The GPCM then is written as

$$P_{jk}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} Z_{jv}(\theta)\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^{c} Z_{jv}(\theta)\right]} \tag{2}$$

351

and

$$Z_{jv}(\theta) = Da_j(\theta - b_{jv}) = Da_j(\theta - b_j + d_v) \ , \tag{3}$$

where

D is a scaling constant that puts the trait (θ) scale in the same metric as the normal ogive model ($D = 1.7$),

a_j is a slope parameter,

b_{jv} is an item-category parameter,

b_j is an item location parameter, and

d_v is a category parameter.

For a binary cognitive item, the item location parameter, b_j, generally is called an item difficulty parameter.

For m_j response categories, only $m_j - 1$ category parameters can be estimated. Any one of the m_j category parameters can be defined as any value, because the term including the parameter is canceled from both the numerator and denominator of the model (Muraki, 1992). Then, d_1 can arbitrarily be defined as 0.

If a Likert scale is used to evoke categorical responses for a set of items, the hypothesis that the set of items shares common category parameters can be tested. If the assumption of a common set of category parameters for the items is not appropriate, a model can be fit in which category parameters differ between items. A step-wise application of these two models was demonstrated by Muraki (1992). The PCM with a constant slope parameter was introduced by Masters (1982). This model is a special case of the GPCM (Equation 2). The comparison between these models also was discussed by Muraki (1992).

For the GPCM, there is an indeterminancy in the set of category parameters and the location parameter. The following constraint, called a location constraint, is imposed on the category parameters within a categorical scale:

$$\sum_{k=2}^{m_j} d_k = 0 \ . \tag{4}$$

The PCM contains the element $Z_{jv}^+(\theta)$; that is,

$$Z_{jk}^+(\theta) = \sum_{v=1}^{k} Z_{jv}(\theta) \ , \tag{5}$$

where $Z_{jv}(\theta)$ is defined by Equation 3. The sum of $Z_{jv}(\theta)$ in Equation 5 can be written as

$$Z_{jk}^+(\theta) = Da_j \left[k(\theta - b_j) + \sum_{v=1}^{k} d_v \right] \ , \tag{6}$$

and the model can be rewritten as

$$Z_{jk}^+(\theta) = Da_j \left[T_k(\theta - b_j) + K_k \right] \ . \tag{7}$$

Andrich (1978) called T_k the scoring function and K_k the category coefficient.

The log-odds of the model probabilities, $P_{j,k-1}(\theta)$ and $P_{jk}(\theta)$, can be expressed as

$$\lambda_{j,k|k-1} = Da_j \left[(T_k - T_{k-1})(\theta - b_j) + d_k \right] \ . \tag{8}$$

Equation 8 shows that the log-odds becomes a monotonically increasing function of θ, given certain fixed values of b_j and d_k and a certain positive value of a_j, only when the incremental change in the

scoring function $(T_k > T_{k-1})$ is used for successive categorical responses. The higher θ value a person has, the more likely he/she responds to upper categories. In other words, the PCM becomes a model for ordered categorical responses only when the scoring function is increasing; that is, $T_k > T_{k-1}$, for any k and $a_j > 0$.

For Masters' (1982) PCM, the scoring function T_k is a linear integer scoring function; that is, $\mathbf{T} = (1, 2, 3, \ldots, m_j)$. If the linear integer scoring function is used, the item-category response functions (ICRFs) of $P_{j,k-1}(\theta)$ and $P_{jk}(\theta)$ intersect at the point b_{jk} on the θ scale. When a person's θ is greater than b_{jk}, the person is more likely to select the kth category than the $k - 1$st category. In this case, the log-odds in Equation 8 is positive. The log-odds increases as θ increases. When a person's θ is less than b_{jk}, the person is more likely to select the $k - 1$st category than the kth category. In this case, the log-odds is negative. The odds of selecting the $k - 1$st category compared to the kth category increases as θ decreases.

The GPCM in Equation 2 can be rewritten, using the scoring function and the category coefficient, as

$$P_{jk}(\theta) = \frac{\exp\{Da_j[T_k(\theta - b_j) + K_k]\}}{\sum\limits_{c=1}^{m_j} \exp\{Da_j[T_c(\theta - b_j) + K_c]\}} \; . \tag{9}$$

The model expressed by Equation 9 is similar to the nominal response model (Bock, 1972). The relationship to the GPCM was first discussed by Samejima (1972) and further elaborated by Thissen & Steinberg (1986).

The scoring function provides a convenient notation for collapsing or recoding categorical responses. For example, if the number of categorical responses of an item is five, then a scoring function $\mathbf{T}$ can be specified as $\mathbf{T} = (1, 2, 3, 4, 5)$. If the original response categories are collapsed by combining the first and second categories into one category, the modified scoring function $\mathbf{T}'$ can be written as $\mathbf{T}' = (1, 1, 2, 3, 4)$. If this modification of the response categories is recoded by treating the original fourth category as the fifth and the original fifth as the fourth, the scoring function can be further modified to $\mathbf{T}'' = (1, 1, 2, 4, 3)$.

Information Functions

The item information function (IIF), $I_j(\theta)$, represents the information contributed by a specific item j across the θ range. Item information for polytomous item response models was proposed by Samejima (1974) as

$$I_j(\theta) = \sum_{k=1}^{m_j} P_{jk}(\theta)\left[-\frac{\partial^2}{\partial\theta^2} \ln P_{jk}(\theta)\right] = \sum_{k=1}^{m_j} P_{jk}(\theta)\left\{\frac{\left[\frac{\partial}{\partial\theta}P_{jk}(\theta)\right]^2}{P_{jk}(\theta)} - \frac{\frac{\partial}{\partial\theta^2}P_{jk}(\theta)}{P_{jk}(\theta)}\right\}$$

$$= D^2 a_j^2 \sum_{k=1}^{m_j} P_{jk}(\theta)\left\{\sum_{c=1}^{m_j} T_c^2 P_{jc}(\theta) - \left[\sum_{c=1}^{m_j} T_c P_{jc}(\theta)\right]^2\right\}$$

$$= D^2 a_j^2 \left\{\sum_{c=1}^{m_j} T_c^2 P_{jc}(\theta) - \left[\sum_{c=1}^{m_j} T_c P_{jc}(\theta)\right]^2\right\} = D^2 a_j^2 \sum_{c=1}^{m_j} [T_c - \bar{T}_j(\theta)]^2 P_{jc}(\theta) \; , \tag{10}$$

where

$$\bar{T}_j(\theta) = \sum_{c=1}^{m_j} T_c P_{jc}(\theta) \; . \tag{11}$$

For the case of dichotomous item responses, Equation 10 can be simplified to

$$I_j(\theta) = D^2 a_j^2 (T_1 - T_2)^2 P_{j1}(\theta) P_{j2}(\theta) \; , \tag{12}$$

where $P_{j2}(\theta) = 1 - P_{j1}(\theta)$.

$\bar{T}_j(\theta)$ in Equation 11 is the item response function (IRF) for a polytomously-scored item with a given scoring function (T_k, $k = 1, 2, \ldots, m_j$) and it can be viewed as a regression of item score on a latent trait scale (Lord, 1980). For a dichotomous item response theory model, the IRF coincides with the ICRF. In the GPCM, the IRF is the conditional mean of item scores at a given θ, and the information is its conditional variance multiplied by $(Da)^2$.

Bock (1972) proposed defining the information due to the response in category k of item j as a partition of the item information due to that category; that is,

$$I_{jk}(\theta) = P_{jk}(\theta) I_j(\theta) . \tag{13}$$

Equation 13 may be called the item-category information function (ICIF). The ICIFs of $I_{j,k-1}(\theta)$ and $I_{jk}(\theta)$ intersect at the point b_{jk} on the θ scale.

The IIF also can be expressed as the summation of the ICIF:

$$I_j(\theta) = \sum_{k=1}^{m_j} I_{jk}(\theta) . \tag{14}$$

Finally, the test information function is defined as the summation of IIFs:

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta) . \tag{15}$$

Item-Category Response and Information Functions

Plotting the ICRFs and their information functions is essential for interpretation of the parameters of polytomous item response models. Figure 1 shows plots of ICRFs of the GPCM with three categorical responses, where $a_j = 1$ and $\mathbf{T} = (1, 2, 3)$. The items differ only in the values of the item-category parameters, $\mathbf{b}_j = (b_{j1}, b_{j2}, \ldots, b_{jm_j})$. Figures 1a–1e show Items 1–5 with $\mathbf{b}_1 = (0, -2, 4)$, $\mathbf{b}_2 = (0, -2, 2)$, $\mathbf{b}_3 = (0, -2, 0)$, $\mathbf{b}_4 = (0, -2, -2)$, and $\mathbf{b}_5 = (0, -2, -4)$, respectively.

As shown in Figures 1a–1e, the third ICRF, $P_{j3}(\theta)$, shifts to the left along the θ scale as the item-category parameter b_{jk} changes from 4.0 to −4.0, because the intersection of the second and third ICRFs, $P_{j2}(\theta)$ and $P_{j3}(\theta)$, moves from 4.0 to −4.0. Because the third category probability, P_{j3}, becomes dominant over the θ axis from −6.0 to 6.0, the expected frequency of the third categorical response increases, assuming the latent trait is distributed normally with mean of 0.0. As the third category probability moves from right to left, the second category probability is pushed downward. The expected frequency of the second category decreases as the distance between the second and third item-category parameters, b_{j2} and b_{j3}, decreases. b_{j2} and b_{j3} are equal for Item 4, and then the order of these parameters is interchanged for Item 5. As shown in Figure 1e, the polytomous responses for Item 5 become essentially dichotomous item responses because the second category becomes a category that is not likely to occur as a response.

ICIFs and IIFs of these items are shown in Figure 2. The item information of polytomous item responses is not necessarily unimodal like that of dichotomous item responses. If the distance between two adjacent item-category parameters is large, such as b_{12} and b_{13} in Figure 2a, the information becomes lower in the middle range of θ. The loss of information over the middle range of θ becomes less noticeable as the distance between these parameters decreases, as shown in Figure 2b. When the distance between b_{j2} and b_{j3} is 2.0, the IIF looks unimodal, as shown in Figure 2c. This item is the most desirable if the item is designed to cover a wide range of θ for a group of persons whose θ levels are assumed to be normally distributed. If the order of these parameters is interchanged,
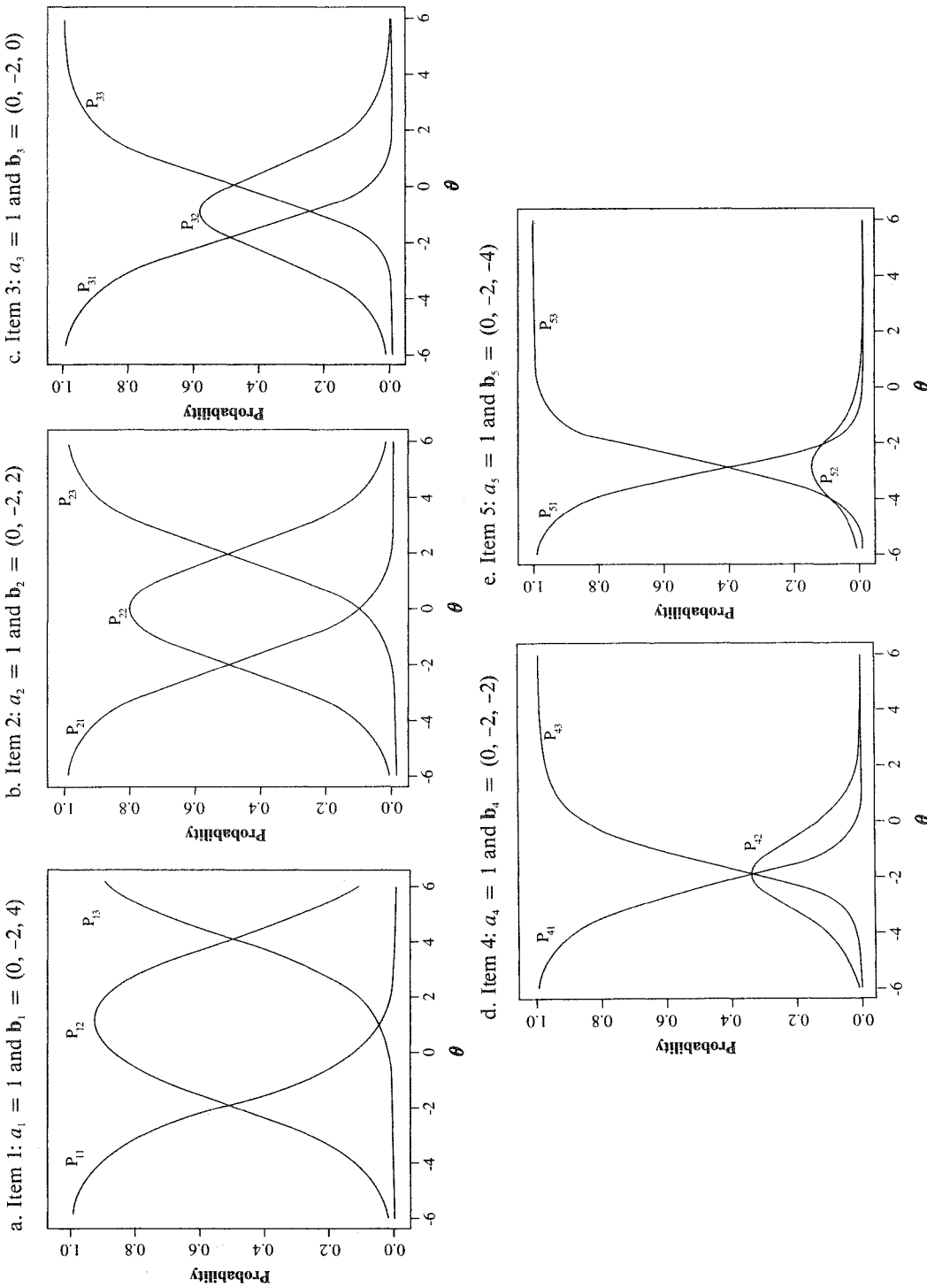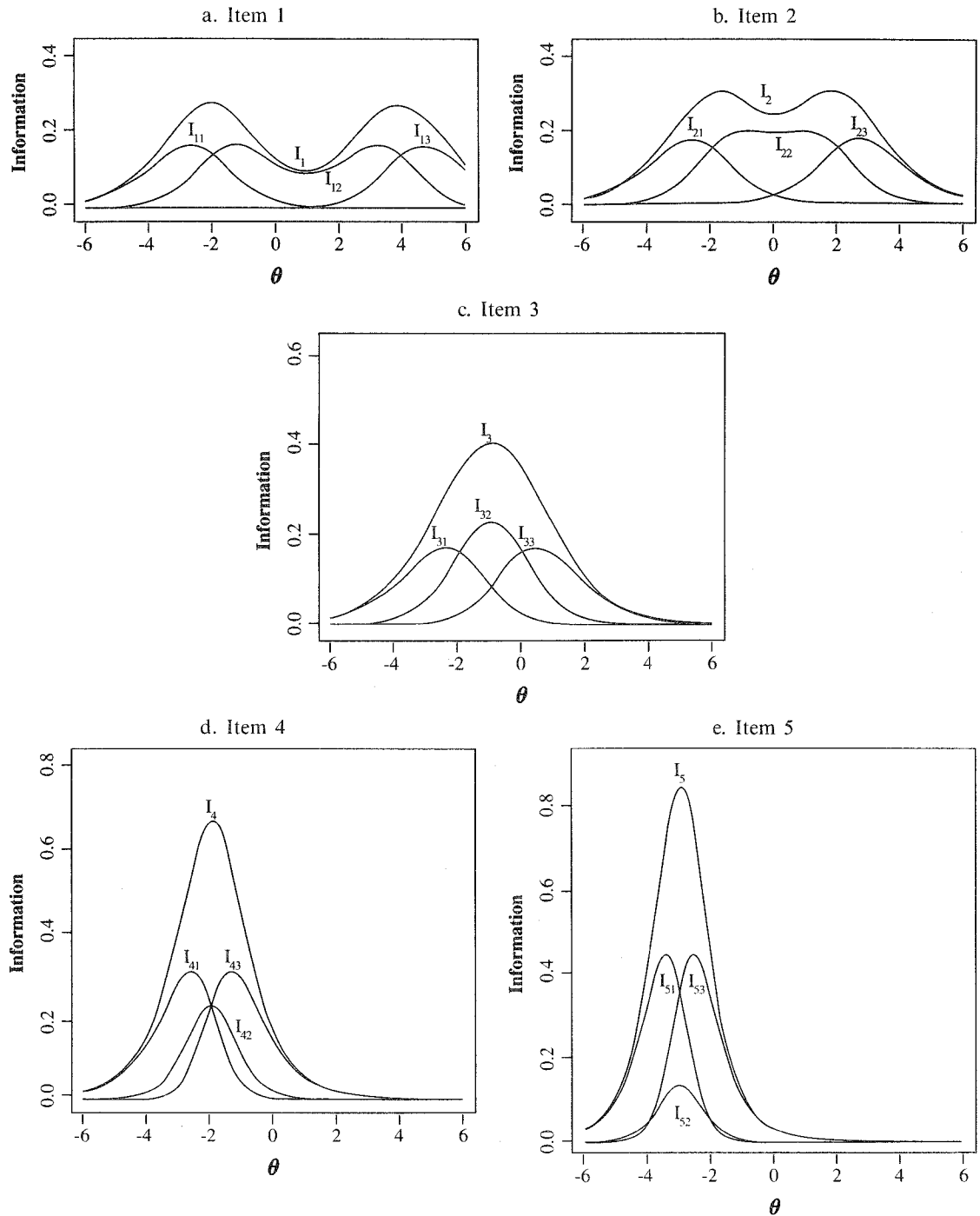
Figure 1
ICRFs of the PCM

a. Item 1: $a_1 = 1$ and $\mathbf{b}_1 = (0, -2, 4)$

b. Item 2: $a_2 = 1$ and $\mathbf{b}_2 = (0, -2, 2)$

c. Item 3: $a_3 = 1$ and $\mathbf{b}_3 = (0, -2, 0)$

d. Item 4: $a_4 = 1$ and $\mathbf{b}_4 = (0, -2, -2)$

e. Item 5: $a_5 = 1$ and $\mathbf{b}_5 = (0, -2, -4)$

Figure 2
IIFs and ICIFs

a. Item 1

b. Item 2

c. Item 3

d. Item 4

e. Item 5

the item information peaks over a very short range of low θ values. Because both item-category parameter values are low, $\mathbf{b}_5 = (0, -2, -4)$, information at the higher θ levels is essentially lost. As shown in Figure 2e, the shape of the IIF resembles that of dichotomous item responses.

Item Response Functions

The first derivative of the IRF with respect to θ is

$$\frac{\partial \bar{T}_j(\theta)}{\partial \theta} = Da_j \sum_{c=1}^{m_j} T_c[T_c - \bar{T}_j(\theta)]P_{jc}(\theta) = Da_j \sum_{c=1}^{m_j} [T_c - \bar{T}_j(\theta)]^2 P_{jc}(\theta)$$

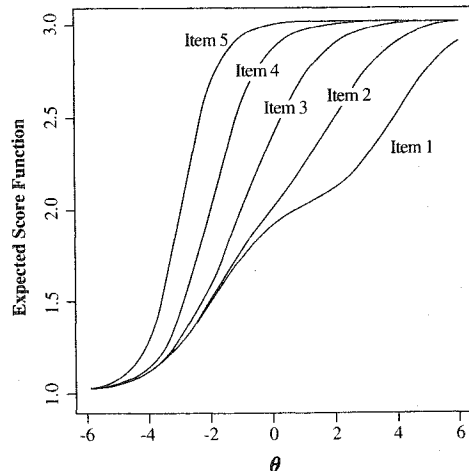$$= \frac{I_j(\theta)}{Da_j} > 0 . \tag{16}$$

Because the first derivative of the IRF with respect to θ in Equation 16 is always positive unless the slope parameter is negative, the IRF is a strictly increasing function of θ. The IRFs of Items 1–5 are plotted in Figure 3. The IRF becomes steeper if the slope parameter value or the item information increases. The slope of the IRF is largest at the point of θ where the item information is a maximum. This is the reason that the IRFs of the five items are ordered along the θ axis in Figure 3. Item 1 appears to have two extreme slopes because its item information has two modes (Figure 2a).

Unlike the category parameters of the graded response model (Muraki, 1990; Samejima, 1969, 1972), the item-category, b_{jk}, or category parameters, d_k, are not necessarily ordered within an item. However, if the order of those parameters is interchanged, as in Item 5, the category is depressed and the category becomes useless in terms of the contribution of the item-category information to the total item information. Such items are usually undesirable.

Collapsing and Reordering Categorical Responses

The ICIFs can be used to investigate the appropriateness of collapsing or recoding categorical responses. One item from the 1990 mathematics cross-sectional assessment and two items (life science and physical science) from the 1990 science cross-sectional assessment of the National Assessment
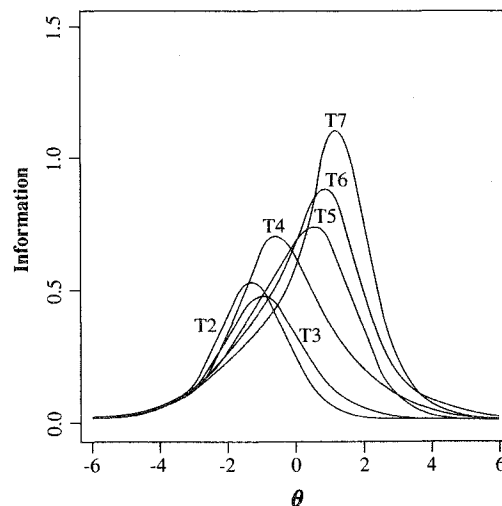
Figure 3
IRFs of Five Items

of Educational Progress (NAEP; Johnson & Allen, 1992) were used to illustrate the behavior of ICIFs with respect to the collapsing and recoding of categorical responses. Item parameters were calibrated using the PARSCALE program (Muraki & Bock, 1991).

Mathematics item. For a mathematics item with seven categories, the scoring function was denoted $T_7 = (1, 2, 3, 4, 5, 6, 7)$. The item parameters of this item were estimated with other items in the assessment. Then, the sixth and seventh categorical responses were combined and all item parameters were estimated again because collapsing or reordering response categories of a subset of items affects the estimates of other items in the PCM (Jansen & Roksam, 1986). The scoring function after collapsing was denoted $T_6 = (1, 2, 3, 4, 5, 6, 6)$. The collapsing process was continued until only two categories remained. The scoring function of this collapsing was denoted $T_2 = (1, 2, 2, 2, 2, 2, 2)$. After each collapsing, the model parameters were estimated and the IIFs were computed.

The IIFs for the six scoring functions are shown in Figure 4. Because the highest category was collapsed with the adjacent categories in sequence, the peak of the IIF moves to the left along the θ scale. In other words, by collapsing higher categorical responses, the information about persons with higher θ values decreased. At the same time, the maximum information decreased, except when changing from T_3 to T_2.

Life science item. A life science item with six categorical responses was administered to two age groups. The ICIFs and IIF based on the item parameters estimated with the scoring function $T = (1, 2, 3, 4, 5, 6)$ are plotted in Figure 5a. The means, v_k, of the total score distribution for those examinees responding in category k $(k = 1, 2, \ldots, 6)$ of the item were computed. These means were $v = (6.3, 7.7, 7.9, 8.3, 8.9, 10.2)$ and $v = (8.5, 8.8, 9.4, 9.4, 10.5, 11.9)$ for the first and second age groups, respectively. The means for the first and second categories for the second age group and the means of the third and fourth categories for both age groups were very close. Thus, these categories did not seem to be differentiated. Allen (1992) then combined these categories and formed four categorical responses. The scoring function for this collapsing was denoted $T = (1, 1, 2, 2, 3, 4)$, and the ICIFs and IIF based on this scoring function are plotted in Figure 5b. The peak of the
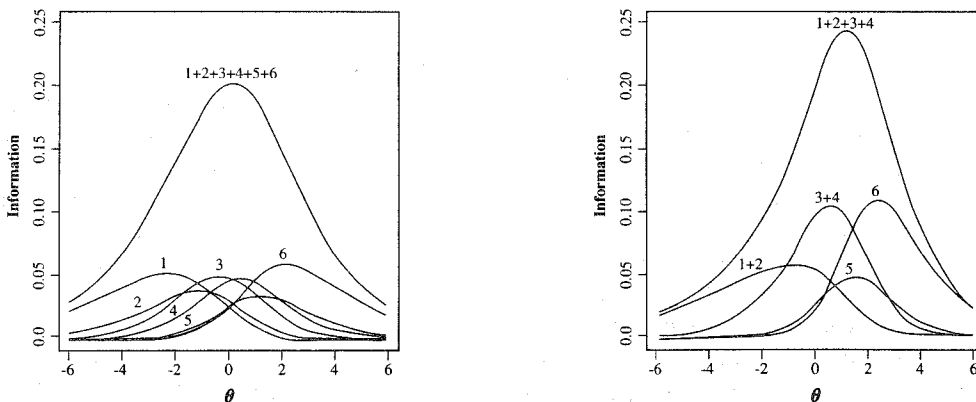
Figure 4
IIF of a Math Item [$T_7 = (1, 2, 3, 4, 5, 6, 7)$; $T_6 = (1, 2, 3, 4, 5, 6, 6)$; $T_5 = (1, 2, 3, 4, 5, 5, 5)$; $T_4 = (1, 2, 3, 4, 4, 4, 4)$; $T_3 = (1, 2, 3, 3, 3, 3, 3)$; $T_2 = (1, 2, 2, 2, 2, 2, 2)$]

Figure 5
IIF and ICIFs for a Life Science Item

a. Before Collapsing Categories [**T** = (1, 2, 3, 4, 5, 6);
   a = .17, **b** = (0.0, -.66, -1.72, .31, 1.93, -.18)]

b. After Collapsing Categories [**T** = (1, 1, 2, 2, 3, 4);
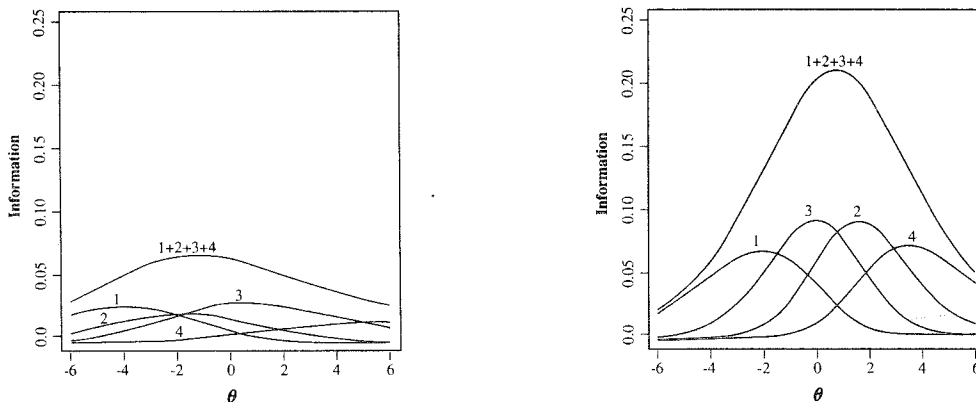 a = .28, **b** = (0.0, -1.2, -2.76, .21)]

IIF becomes higher after the collapsing, and the amount of information increases over the θ scale, except at its lower end which may be explained by the effect of collapsing lower categories.

   Physical science item.   A physical science item with four categories also was administered to two age groups. The ICIFs and IIF based on the item parameters estimated with the scoring function, **T** = (1, 2, 3, 4), are plotted in Figure 6a. The mean vectors for the two age groups were **v** = (4.8, 6.4, 6.2, 7.5) and **v** = (9.3, 10.9, 10.6, 11.5). These statistics suggest that the original codings of the second and third categories were inappropriate (Allen, 1992), because the mean scores for these two categories were reversed from expectations. Thus, the orders of the second and third categories were reversed, and the scoring function was denoted **T'** = (1, 3, 2, 4). The ICIFs and IIF based on this scoring function are plotted in Figure 6b. The information increased considerably over the θ range

Figure 6
IIF and ICIFs for a Physical Science Item

a. Before Recoding Categories [**T** = (1, 2, 3, 4);
   a = .17, **b** = (0.0, -2.11, -1.76, 5.19)]

b. After Recoding Categories [**T** = (1, 3, 2, 4);
   a = .32, **b** = (0.0, -1.56, -.60, 2.95)]

−4.0 to 4.0. The χ^2 fit statistic also improved from 370.002 with 87 degrees of freedom (df) to 151.249 with 82 df. The difference, 218.753, with 5 df is a significant improvement for model fit.

Analysis of the NAEP Writing Assessment

Method

The first NAEP writing assessment was conducted in 1984 (Grima & Johnson, 1992). Since then, four NAEP writing assessments have been conducted to assess the trend of students' writing performance over the years 1984, 1988, 1990, and 1992. In the trend assessment, nationally representative samples of students in grades 4, 8, and 11 responded to a series of writing items (or tasks).

A set of 12 writing items was prepared to examine students' abilities to engage in three types of writing: informative, persuasive, and imaginative. The 12 items were administered in a balanced incomplete block design with six of the 12 items presented to each grade. Some writing items were unique to a specific grade, and others were used as linking items. Students' writing samples were scored by trained readers on the basis of the students' success in accomplishing the specific purpose of each writing item (as measured by primary trait scoring), their relative writing fluency (as measured by holistic scoring), and their mastery of the conventions of writing English (as measured by their spelling, punctuation, and grammar). For this study, only the primary trait scores of the 1988 and 1990 assessments were analyzed.

For two writing items, a four-point scoring scale (unsatisfactory through adequate) was used to evaluate the students' writing samples. For the remaining four writing items, a five-point scoring scale (unsatisfactory through elaborate) was used. Omitted responses were not rated; rather, omitted responses were treated as the lowest categorical response (unsatisfactory).

Item parameters were calibrated based on the combined response data of the 1988 and 1990 assessment. Because it could not be assumed that the latent trait distributions of these assessments were unchanged, a separate normal prior distribution was used for each assessment year. 41 quadrature points were used for both prior distributions. After each EM cycle, the means and standard deviations were computed for these distributions and the weighted mean and standard deviation of the combined distribution of these priors were adjusted to 0.0 and 1.0, respectively. Intermediate estimated values also were adjusted accordingly.

The sample sizes for the 1988 and 1990 assessments were 4,878 and 5,606, respectively. 46 EM cycles were needed for convergence, using a criterion of .001. Because there were no responses in the fifth category of Item 2 in the 1990 assessment, the fourth and fifth categories of Item 2 were combined for both years. Thus, Item 2 was treated as a four-category item.

Results

The estimated parameter values for age 9 are presented in Table 1. The IIFs of these six items are plotted in Figure 7. The test information function is plotted in Figure 8.

Because the slope parameter of Item 1 was low (.307), the IIF (Figure 7a) was relatively flat over the θ scale. In addition to a low slope parameter, Item 3 also had widely-dispersed category parameters. Because of this combination, when compared to the other items, the IIF for this item (Figure 7c) was the lowest of the six items for all θ values. The slope parameter of Item 4 (.706) was not extremely low compared to Items 1 and 3, but the category parameters were considerably dispersed. Consequently, the amount of information increased over the θ scale, but the information curve was relatively flat (Figure 7d). The best items among the six were Items 5 and 6 (Figures 7e and 7f). The amount of information for Item 5 was high for the lower end of the θ scale because the location

Table 1
Item Parameters and Item-Category Parameters
of NAEP Writing Items (1988 and 1990, Age 9)

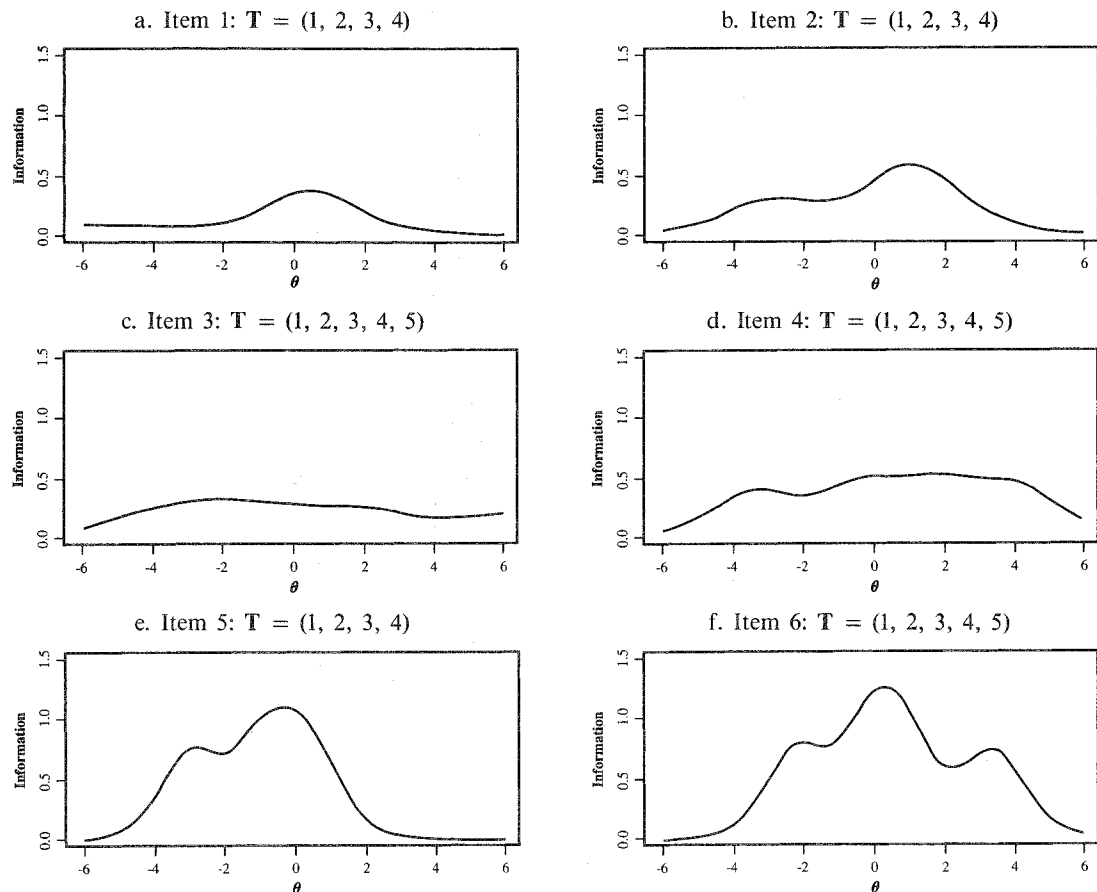| Item	Item Parameter a_j	Item Parameter b_j	d_k	b_{jk}
1	.370	−1.362	4.015	−5.377
			−4.835	3.473
			.819	−2.181
2	.617	−.288	2.744	−3.032
			−.747	.459
			−1.997	1.709
3	.507	1.053	4.628	−3.575
			2.083	−1.030
			−1.053	2.106
			−5.659	6.712
4	.706	.595	4.059	−3.464
			.983	−.388
			−1.351	1.946
			−3.691	4.286
5	.969	−1.114	1.844	−2.958
			−.252	−.862
			−1.592	.478
6	.981	.522	2.705	−2.183
			.677	−.156
			−.441	.963
			−2.942	3.463

parameter of this item was low. Compared to Item 5, the information function of Item 6 was more symmetric and reasonably high between $\theta = -3$ to 3. The test information plot in Figure 8 shows that this set of six items produced a reasonable amount of information between $\theta = -3$ to 3.

Conclusions

Because of the complex relationships among the parameters of the PCM, analysis of the plots of the ICRFs and IRF should be a routine step for evaluation of these kinds of test items. Computing and plotting the ICIF and IIFs based on the PCM is also an essential procedure of item analyses. Information functions, along with other conventional item statistics, provide valuable information about how to collapse or reorder categorical responses. Increasing the number of categorical responses does not uniformly increase the information for θ levels over the entire θ range, although polytomous response items generally cover a wider range of the θ scale than dichotomous items. Assembling polytomous items into a desirable test also is facilitated by using the information functions. Careful investigation of each test item in terms of its information function leads to desirable test construction methodology for polytomously-scored items.

Applications of the PCM to polytomous items discussed here are only an initial endeavor for this new area in psychometrics. Jansen & Roskam (1986) compared the PCM with the graded response model and concluded that the latter model was more desirable for rating data because of its invariant property of the latent trait scale after collapsing categories. This undesirable effect specific to the PCM was minimized by collapsing only a single item in each assessment and by recalibrating items by the marginal maximum likelihood estimation method with a fixed normal prior distribution of

Figure 7
IIFs for Six Items

a. Item 1: $T = (1, 2, 3, 4)$

b. Item 2: $T = (1, 2, 3, 4)$

c. Item 3: $T = (1, 2, 3, 4, 5)$

d. Item 4: $T = (1, 2, 3, 4, 5)$

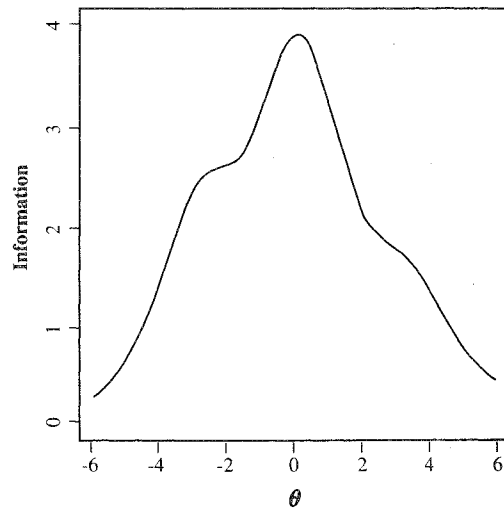e. Item 5: $T = (1, 2, 3, 4)$

f. Item 6: $T = (1, 2, 3, 4, 5)$

θ (Muraki, 1992). However, further investigation of the PCM and its comparison with the graded response model is needed.

References

Allen, N. L. (1992, February). *An application of BILOG/PARSCALE to the 1990 science cross-sectional items.* Paper presented at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Alexandria, VA.

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

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29–51.

Grima, A. M., & Johnson, E. G. (1992). Data analysis for the writing assessment. In E. G. Johnson & N. L. Allen (Eds.), *The NAEP 1990 technical report* (Technical Rep. No. 21-TR-20). Princeton NJ: Educational Testing Service, National Assessment of Educational Progress.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston MA: Kluwer-Nijhoff.

Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 51,* 69–91.

Johnson, E. G., & Allen, N. L. (1992). *The NAEP 1990 technical report* (Technical Rep. No. 21-TR-20). Princeton NJ: Educational Testing Service, National Assessment of Educational Progress.

Figure 8
Test Information Function of the NAEP Writing Items

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

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

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement, 14,* 59–71.

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

Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data* [Computer program]. Chicago IL: Scientific Software.

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

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph,* No. 18.

Samejima, F. (1974). Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika, 39,* 111–121.

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

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago IL: MESA.

Author's Address

Send requests for reprints or further information to Eiji Muraki, Educational Testing Service, Princeton NJ 08541, U.S.A. Internet: emuraki@rosedale.org.