PSYCHOMETRIKA—VOL. 60, NO. 4, 549–572
DECEMBER 1995

ACCELERATION MODEL IN THE HETEROGENEOUS CASE OF THE GENERAL GRADED RESPONSE MODEL

FUMIKO SAMEJIMA

THE UNIVERSITY OF TENNESSEE

A new model, called acceleration model, is proposed in the framework of the heterogeneous case of the graded response model, based on processing functions defined for a finite or enumerable number of steps. The model is expected to be useful in cognitive assessment, as well as in more traditional areas of application of latent trait models. Criteria for evaluating models are proposed, and soundness and robustness of the acceleration model are discussed. Graded response models based on individual choice behavior are also discussed, and criticisms on model selection in terms of fitnesses of models to the data are also given.

Key words: latent trait models, graded response models, item response theory, partial credit models, model evaluation, cognitive assessment, processing functions, operating characteristics.

Samejima (1972) has proposed a general theoretical framework of the graded response model, in which the *homogeneous case* is distinguished from the *heterogeneous case*. The *general graded response model* represents a family of mathematical models which deal with *ordered polychotomous categories* in general. These ordered categories include: A, B, C, D and F in the evaluation of students' performance, *strongly disagree*, *disagree*, *agree* and *strongly agree* in a social attitude survey, *partial credit* given in accordance with the individual's degree of attainment toward the solution of a problem, to give some examples.

With a rapid progress of computer technologies, today we can: (a) *program* a well-controlled cognitive experiment in computer *software*, (b) accommodate the software in a number of microcomputers, and (c) have each trained experimenter carry one of the microcomputers and collect data, conducting the cognitive experiment on individuals, as we do in a small-room experimental situation. In this way, we can easily collect data for several hundred individual subjects within a limited amount of time. This is a sample size comparable to those for a paper-and-pencil test in a college campus environment. Thus with such a set of data not only can we observe each individual's behavior intensively just as in a cognitive experiment, but also we can analyze such a *rich* set of data applying psychometric theories, put the results *in perspective*, clarify individual differences, et cetera.

The present paper proposes a mathematical model, called *acceleration model*, in an effort to provide a mathematical model in the framework of general graded response model which is sound and useful in analyzing such intensive cognitive data as well as more traditional psychometric data.

General Graded Response Model

Let θ be the latent trait, or *ability*, which represents a construct hypothesized underneath certain human behavior, and is assumed to take on any real number. Let g

This research was supported by the Office of Naval Research (N00014-90-J-1456).

Requests for reprints should be sent to Fumiko Samejima, Department of Psychology, 310B Austin Peay Bldg., University of Tennessee, Knoxville, TN 37996-0900. E-mail: samejima@psych1.psych.utk.edu

0033-3123/95/1200-94196$00.75/0

denote an item, which is the smallest unit of manifest entity for measuring θ. Let X_g be a graded item response to item g, and x_g $(= 0, 1, \ldots, m_g)$ denote its realization. The *operating characteristic*, $P_{x_g}(\theta)$, of the item score x_g means the conditional probability, given θ, with which the individual of ability θ gets x_g, that is,

$$P_{x_g}(\theta) \equiv \text{Prob}\,[X_g = x_g|\theta],$$

which is assumed to be five times differentiable with respect to θ. For convenience, hereafter, x_g will be used both for a specific discrete response and for the event $X_g = x_g$, and a similar usage is applied for other symbols.

For a set of $n(\geq 1)$ items, a response pattern, denoted by V, indicates a sequence of X_g for $g = 1, 2, \ldots, n$, and its realization, v, can be written as

$$v = \{x_g\}'.$$

It is assumed that *local independence* (Lord & Novick, 1968) holds, so that within any group of individuals all characterized by the same value of ability θ the distributions of the item responses are all independent of each other. Thus the operating characteristic, $P_v(\theta)$, of the response pattern v is defined as

$$P_v(\theta) \equiv \text{Prob}\,[V = v|\theta] = \prod_{x_g \in v} P_{x_g}(\theta), \tag{1}$$

which is also the likelihood function, $L(v|\theta)$, for $V = v$.

Processing Functions

Suppose, for example, that a cognitive process, like *problem solving*, contains a finite or enumerable number of steps. The graded item score x_g should be assigned to the individuals who have successfully completed up to the step x_g but *failed* to complete the step $(x_g + 1)$. Let $M_{x_g}(\theta)$ be the *processing function* of the graded item score x_g, which is the *joint* conditional probability with which the individual completes the step x_g successfully, under the conditions that:

1. the individual's ability level is θ, and
2. the steps up to $(x_g - 1)$ have already been followed and completed successfully.

It is assumed that $M_{x_g}(\theta)$ is either strictly increasing in θ or constant for all θ, for $x_g = 1, 2, \ldots, m_g$. This assumption is reasonable considering that each item has some direct and positive significance to the ability measured. Let $(m_g + 1)$ be the hypothesized graded item score adjacent to and above m_g. Since everyone can at least obtain the item score 0, and no one is able to obtain the item score $(m_g + 1)$, it is reasonable to set

$$M_{x_g}(\theta)\begin{cases} = 1 & \text{for } x_g = 0 \\ = 0 & \text{for } x_g = m_g + 1, \end{cases}$$

for all θ.

Fundamental Framework

Thus the operating characteristic, $P_{x_g}(\theta)$, of the graded item score x_g is given by

$$P_{x_g}(\theta) = \prod_{u \leq x_g} M_u(\theta)[1 - M_{(x_g+1)}(\theta)]. \tag{2}$$

This provides the *fundamental framework* for the general graded response model. Let $P^*_{x_g}(\theta)$ denote the conditional probability with which the individual of ability θ follows and completes the cognitive process successfully up to the step x_g, *or further*. Thus

$$P^*_{x_g}(\theta) = \prod_{u \le x_g} M_u(\theta). \tag{3}$$

This function is called the *cumulative operating characteristic* although cumulation is in the opposite direction. Note that $P^*_{x_g}(\theta)$ becomes the operating characteristic, $P_g(\theta)$, of the positive response to item g, when the graded item score x_g is changed to the binary score, assigning 0 to all scores less than x_g and 1 to those score categories greater than or equal to x_g. From (2) and (3) the operating characteristic, $P_{x_g}(\theta)$, can also be expressed by

$$P_{x_g}(\theta) = P^*_{x_g}(\theta) - P^*_{(x_g+1)}(\theta). \tag{4}$$

It is obvious from (3) that $P^*_{x_g}(\theta)$ is also either strictly increasing in θ or constant for all θ, and assumes unity for $x_g = 0$ and zero for $x_g = m_g + 1$ for the entire range of θ.

The homogeneous case of the graded response model represents a family of models in which $P^*_{x_g}(\theta)$'s for $x_g = 1, 2, \ldots, m_g$ are identical in shape, and these m_g functions are positioned alongside the abscissa in accordance with the item score x_g. The heterogeneous case of the graded response model represents all mathematical models that provide a set of cumulative operating characteristics, $P^*_{x_g}(\theta)$'s, *not all* of which are identical in shape, that is, those which do not belong to the homogeneous case.

The *basic function*, $A_{x_g}(\theta)$, is defined by

$$A_{x_g}(\theta) \equiv \frac{\partial}{\partial\theta} \log P_{x_g}(\theta) = \sum_{u \le x_g} \frac{\partial}{\partial\theta} \log M_u(\theta) + \frac{\partial}{\partial\theta} \log\,[1 - M_{(x_g+1)}(\theta)], \tag{5}$$

which has an important role in computer algorithm for obtaining the maximum likelihood estimate of the individual's ability from his/her response pattern, et cetera (see Samejima, 1969, 1972, 1973b).

The *item response information function* (Samejima, 1973b, 1994) of the graded item response x_g is defined by

$$I_{x_g}(\theta) \equiv -\frac{\partial^2}{\partial\theta^2} \log P_{x_g}(\theta) = -\frac{\partial}{\partial\theta} A_{x_g}(\theta) = \sum_{u \le x_g} -\frac{\partial^2}{\partial\theta^2} \log M_u(\theta)$$
$$-\frac{\partial^2}{\partial\theta^2} \log\,[1 - M_{(x_g+1)}(\theta)]; \tag{6}$$

and the *item information function*, is obtained as its conditional expectation, given θ, so that

$$I_g(\theta) \equiv E[I_{x_g}(\theta) | \theta] = \sum_{x_g} I_{x_g}(\theta) P_{x_g}(\theta), \tag{7}$$

which includes Birnbaum's item information function for the dichotomous test item (Birnbaum, 1968) as a special case. The *response pattern information function* (Samejima, 1973b) is given by

$$I_{\underline{v}}(\theta) \equiv -\frac{\partial^2}{\partial\theta^2} \log P_v(\theta) = \sum_{x_g \in v} \left[-\frac{\partial^2}{\partial\theta^2} \log P_{x_g}(\theta)\right] = \sum_{x_g \in v} I_{x_g}(\theta),$$

and the *test information function* is defined as the conditional expectation of $I_v(\theta)$, given θ, to obtain

$$I(\theta) \equiv E[I_v(\theta)|\theta] = \sum_v I_v(\theta)P_v(\theta) = \sum_{g=1}^{n} I_g(\theta),$$

which includes Birnbaum's test information function on the dichotomous response level as a special case. This has been used as a local measure of accuracy in ability estimation, and its modified formulas (Samejima, 1994) have also been proposed, using the MLE bias function.

Acceleration Model

The *acceleration model* is a model which belongs to the heterogeneous case of the graded response model. It has been built in an effort to provide a model which is useful in cognitive assessment (Samejima, 1995), as well as in more traditional analysis of test data, et cetera.

Consider a situation, such as problem solving, that requires a number of subprocesses before attaining the solution. It is assumed that there is more than one *step* in the whole process which is *observable*. Graded item scores, or partial credits, 1 through m_g, are assigned to the successful completions of these separate observable steps.

The processing function in this model for each x_g (= 1, 2, . . . , m_g) is given by

$$M_{x_g}(\theta) = [\Psi_{x_g}(\theta)]^{\xi_{x_g}}, \tag{8}$$

where ξ_{x_g} (>0) is the step *acceleration* parameter. The acceleration model represents a family of models in which $\Psi_{x_g}(\theta)$ is specified by a strictly increasing, five times differentiable function of θ with zero and unity as its two asymptotes, and the ratio

$$\frac{\dfrac{\partial}{\partial\theta}\log\dfrac{\partial}{\partial\theta}\Psi_{x_g}(\theta)}{\dfrac{\partial}{\partial\theta}\log\Psi_{x_g}(\theta)} = \frac{\Psi_{x_g}(\theta)\dfrac{\partial^2}{\partial\theta^2}\Psi_{x_g}(\theta)}{\left[\dfrac{\partial}{\partial\theta}\Psi_{x_g}(\theta)\right]^2} \tag{9}$$

decreases with θ.

From (3) and (8) the cumulative operating characteristic, $P^*_{x_g}(\theta)$, is given by

$$P^*_{x_g}(\theta) = \prod_{u=0}^{x_g} [\Psi_u(\theta)]^{\xi_u}.$$

We obtain from (2) and (8) the operating characteristic such that

$$P_{x_g}(\theta) = \prod_{u=0}^{x_g} [\Psi_u(\theta)]^{\xi_u}[1 - [\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}}]. \tag{10}$$

From (8), we can write

$$\frac{\partial}{\partial\theta} M_{x_g}(\theta) = \xi_{x_g}\Psi_{x_g}(\theta)^{\xi_{x_g}-1}\frac{\partial}{\partial\theta}\Psi_{x_g}(\theta),$$

and

$$\frac{\partial^2}{\partial\theta^2} M_{x_g}(\theta) = \xi_{x_g} \Psi_{x_g}(\theta)^{\xi_{x_g}-2}\left[(\xi_{x_g} - 1)\left[\frac{\partial}{\partial\theta} \Psi_{x_g}(\theta)\right]^2 + \Psi_{x_g}(\theta) \frac{\partial^2}{\partial\theta^2} \Psi_{x_g}(\theta)\right].$$

Setting the above equal to zero, we obtain

$$\xi_{x_g} = 1 - \left[\frac{\Psi_{x_g}(\theta) \frac{\partial^2}{\partial\theta^2} \Psi_{x_g}(\theta)}{\left[\frac{\partial}{\partial\theta} \Psi_{x_g}(\theta)\right]^2}\right].$$

Thus from (9) the value of θ at which the discrimination power of $M_{x_g}(\theta)$ is maximal increases with ξ_{x_g}.

Let w denote a subprocess, which is the smallest unit in the cognitive process. Thus each step contains one or more w's. Let $\xi_w(>0)$ be the subprocess *acceleration* parameter, and then the step acceleration parameter, ξ_{x_g}, for each of $x_g = 1, 2, \ldots, m_g$ is given as the sum of ξ_w's over all $w \in x_g$. The name, acceleration parameter, comes from the fact that, within each step, separate subprocesses contribute to *accelerate* the value of θ at which the discrimination power is maximal to its ultimate position.

It is assumed that the whole process leading to the solution consists of a finite number of *clusters*, each containing one or more steps, and within each cluster the parameters in $\Psi_{x_g}(\theta)$ are common. Thus, if two or more adjacent x_g's belong to the same cluster, then the parameters in $\Psi_{x_g}(\theta)$ are the same for these x_g's, and, otherwise, at least one of the parameters is different.

Example

It will be worthwhile to give a simple example to clarify the relationship between a step x_g and subprocesses w's. Consider the following problem solving.

Prove that

$$\int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx = [2\pi]^{1/2}.$$

If this question is presented in a computerized test and software has been prepared so that the examinee's performance is appropriately recorded, then we may be able to consider nine observable steps in this particular problem solving as shown below.

Step 1:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx = 2\int_{0}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx > 0.$$

Step 2:

$$\int_{0}^{\infty} \exp\left[-\frac{a^2}{2}\right] \int_{0}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx\, da = \int_{0}^{\infty} \exp\left[-\frac{a^2}{2}\right] da \int_{0}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx$$

$$= \left[\int_{0}^{\infty} \exp\left[-\frac{x^2}{2}\right] dx\right]^2.$$

Step 3: Setting $y = x/a$ $(a > 0)$,

$$\int_0^\infty \exp\left[-\frac{x^2}{2}\right] dx = \int_0^\infty \exp\left[-\frac{a^2y^2}{2}\right] \frac{dx}{dy}\, dy = \int_0^\infty \exp\left[-\frac{a^2y^2}{2}\right] a\, dy.$$

Step 4: Substituting Step 3 into Step 2,

$$\int_0^\infty \exp\left[-\frac{a^2}{2}\right] \int_0^\infty \exp\left[-\frac{x^2}{2}\right] dx\, da = \int_0^\infty \exp\left[-\frac{a^2}{2}\right] \int_0^\infty \exp\left[-\frac{a^2y^2}{2}\right] a\, dy\, da$$

$$= \int_0^\infty \int_0^\infty \exp\left[-\frac{a^2(1+y^2)}{2}\right] a\, da\, dy.$$

Step 5:

$$\int_0^\infty \exp\left[-\frac{a^2(1+y^2)}{2}\right] a\, da = -\frac{1}{1+y^2} \exp\left[-\frac{a^2(1+y^2)}{2}\right]\Bigg|_0^\infty = \frac{1}{1+y^2}.$$

Step 6: Substituting Step 5 into Step 4,

$$\int_0^\infty \exp\left[-\frac{a^2}{2}\right] \int_0^\infty \exp\left[-\frac{x^2}{2}\right] dx\, da = \int_0^\infty \frac{1}{1+y^2}\, dy = \tan^{-1} y|_0^\infty = \frac{\pi}{2}.$$

Step 7: From Steps 2 and 6,

$$\left[\int_0^\infty \exp\left[-\frac{x^2}{2}\right] dx\right]^2 = \frac{\pi}{2}.$$

Step 8: Taking the square root of each side of Step 7,

$$\int_0^\infty \exp\left[-\frac{x^2}{2}\right] dx = \left[\frac{\pi}{2}\right]^{1/2}.$$

Step 9: Substituting Step 8 into Step 1,

$$\int_{-\infty}^\infty \exp\left[-\frac{x^2}{2}\right] dx = 2 \times \left[\frac{\pi}{2}\right]^{1/2} = [2\pi]^{1/2}.$$

If two or more steps, or sequences of steps, are reversible in order, we shall say they are *parallel*, as distinct from *serial steps*, whose order cannot be changed. In this example, Step 2 and the sequence of Steps 3 through 6 are parallel.

Each of the above steps contains more than one subprocess. For example, Step 1 includes four subprocesses, that is, (a) realizing that $\exp[x^2/2]$ is symmetric at $x = 0$, (b) thus rewriting the original integration as two times the second integration, (c) realizing that $\exp[-x^2/2]$ is positive for $0 < x < \infty$, and (d) thus the result of the integration is greater than 0.

Note that in Step 8 the *sequential order* between the two subprocesses is arbitrary, that is, the square root of whichever side is evaluated first the step will be successfully completed. If there are two or more subprocesses within a step whose sequential order is arbitrary, then these subprocesses are said to be *parallel*, as distinct from *serial* subprocesses. It is assumed that for any number of parallel subprocesses the subprocess acceleration parameters are invariant across shifts of the positions of the subprocesses in the sequence. Thus the step acceleration parameter, ξ_{x_g} ($= \xi_{w_x g_1} + \xi_{w_x g_2} + \cdots$), will be unchanged regardless of the sequential order of these parallel subprocesses.

It will be safer to treat each step as one which belongs to its own cluster because of slightly different natures of these tasks, although the results may indicate that some successive steps belong to one cluster.

A Specific Model

As a *specific* model that belongs to this family of acceleration model, consider one in which $\Psi_{x_g}(\theta)$ is given by the logistic distribution function, such that

$$\Psi_{x_g}(\theta) = \frac{1}{1 + \exp\left[-D\alpha_{x_g}(\theta - \beta_{x_g})\right]}, \tag{11}$$

where $D = 1.7$, and $\alpha_{x_g}(>0)$ and β_{x_g} are the discrimination and location parameters, respectively. It is obvious from (11) that the ratio given by (9) becomes

$$\frac{\frac{\partial}{\partial\theta} \log \frac{\partial}{\partial\theta} \Psi_{x_g}(\theta)}{\frac{\partial}{\partial\theta} \log \Psi_{x_g}(\theta)} = 1 - \left[\frac{\Psi_{x_g}(\theta)}{1 - \Psi_{x_g}(\theta)}\right],$$

which is a strictly decreasing in θ, and, therefore, (11) satisfies the condition for $\Psi_{x_g}(\theta)$. Thus we have for the processing function

$$M_{x_g}(\theta) = \frac{1}{\left[1 + \exp\left[-D\alpha_{x_g}(\theta - \beta_{x_g})\right]\right]^{\xi_{x_g}}}, \tag{12}$$

and the first partial derivative of the processing function becomes

$$\frac{\partial}{\partial\theta} M_{x_g}(\theta) = \xi_{x_g} D\alpha_{x_g}[\Psi_{x_g}(\theta)]^{\xi_{x_g}}[1 - \Psi_{x_g}(\theta)] > 0, \tag{13}$$

and the second partial derivative is obtained from (13) such that

$$\frac{\partial^2}{\partial\theta^2} M_{x_g}(\theta) = \xi_{x_g} D^2\alpha_{x_g}^2[\Psi_{x_g}(\theta)]^{\xi_{x_g}}[1 - \Psi_{x_g}(\theta)][\xi_{x_g}\{1 - \Psi_{x_g}(\theta)\} - \Psi_{x_g}(\theta)], \tag{14}$$

respectively.

It should be noted that, in (12), the location parameter, β_{x_g}, does *not* necessarily increase with x_g. For instance, in the example of problem solving, Step 2 may be the most critical step, and the location parameter, β_2, is likely to be substantially *higher* than β_3 and β_4. If the tasks involved in the sequential steps leading to the problem solution become progressively more difficult, as is observed in some problem solvings, however, it is likely that β_{x_g} increases with x_g.

The basic function, $A_{x_g}(\theta)$, in this model is obtained from (5) and (12), so that

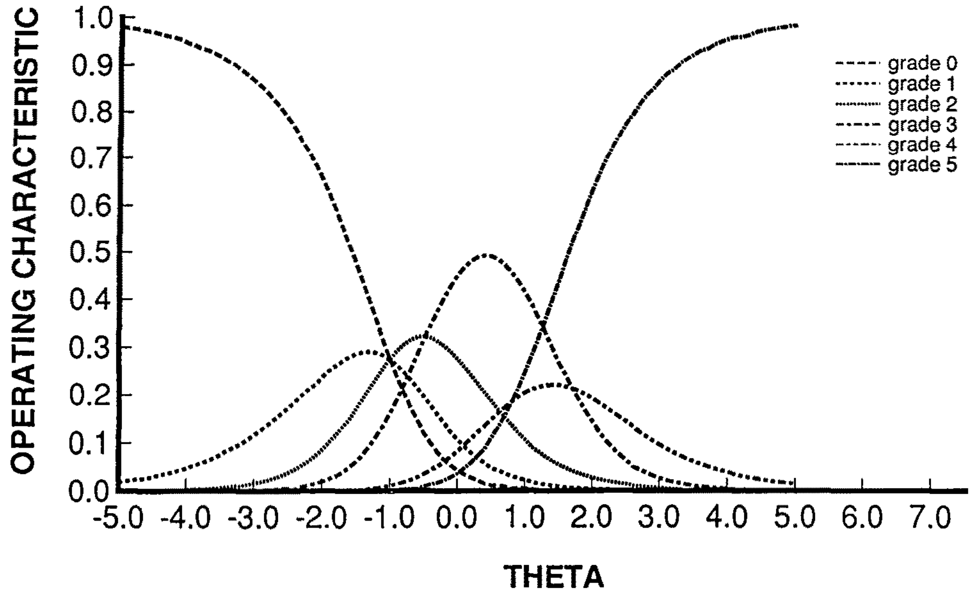

FIGURE 1.
Example of a set of operating characteristics of six steps in the acceleration model.

$$A_{x_g}(\theta) = D\left[\sum_{u \le x_g} \xi_u \alpha_u [1 - \Psi_u(\theta)] - \xi_{x_g+1}\alpha_{x_g+1} \frac{[\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}}[1 - \Psi_{(x_g+1)}(\theta)]}{1 - [\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}}}\right], \tag{15}$$

for $x_g = 1, 2, \ldots, m_g - 1$, and for $x_g = 0$ and $x_g = m_g$ the first term and the second term on the right hand side of (15) disappear, respectively. Item response information function, $I_{x_g}(\theta)$, for this specific model is obtained from (6) and (15) by

$$I_{x_g}(\theta) = D^2\left[\sum_{u \le x_g} \xi_u \alpha_u^2 \Psi_u(\theta)\{1 - \Psi_u(\theta)\} + \xi_{x_g+1}\alpha_{x_g+1}^2[\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}} \cdot \left\{\xi_{x_g+1}\left[\frac{1 - \Psi_{(x_g+1)}(\theta)}{1 - [\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}}}\right]^2 - \Psi_{(x_g+1)}(\theta)\left[\frac{1 - \Psi_{(x_g+1)}(\theta)}{1 - [\Psi_{(x_g+1)}(\theta)]^{\xi_{x_g+1}}}\right]\right\}\right], \tag{16}$$

which assumes positive values for the entire range of θ for $x_g = 0, 1, \ldots, m_g$. Substituting (10) and (16) into (7), the item information function is obtained.

Figure 1 illustrates the six operating characteristics by a solid line, with $m_g = 5$ and the parameters $\alpha_{x_g} = 1.36517, 1.03244, 0.87524, 1.09083, 0.58824$, $\beta_{x_g} = -0.94260, -0.76985, 0.03941, 1.35406, 0.80000$, and $\xi_{x_g} = 0.41972, 0.51741, 0.54196, 0.60004, 1.00000$, for $x_g = 1, 2, 3, 4, 5$.

θ_{dmax} at Which the Processing Function is Most Discriminating

Setting (14) equal to zero, we obtain

$$\theta_{dmax} = \Psi_{x_g}^{-1}\left[\frac{\xi_{x_g}}{1 + \xi_{x_g}}\right], \tag{17}$$

where θ_{dmax} indicates the value of θ at which the processing function $M_{x_g}(\theta)$ is steepest, or most discriminating. It is obvious from (17) that θ_{dmax} is a strictly increasing function of ξ_{x_g}, and

$$\begin{cases} \theta_{dmax} < \beta_{x_g} & \text{if } \xi_{x_g} < 1 \\ \theta_{dmax} = \beta_{x_g} & \text{if } \xi_{x_g} = 1 \\ \theta_{dmax} > \beta_{x_g} & \text{if } \xi_{x_g} > 1. \end{cases}$$

Note that the same set of relationships will hold if we replace ξ_{x_g} by ξ_w in (17). From this we can say that within each step separate subprocesses contribute to *accelerate* θ_{dmax} to its ultimate position.

Parameter Estimation

Parameter estimation in this specific model can be done using the following method.

1. Use a nonparametric estimation method like Levine's (1984) or Samejima's (1983, 1993, 1994, in press), and estimate the operating characteristics, $P_{x_g}(\theta)$'s.
2. Tentatively parameterize the results using a very general semiparametric method, such as Ramsay and Wong's (1993). A strength of this method is that the fit is considered not only for the function in question, but also for its first derivative.
3. From these results obtain the estimated processing function $\hat{M}_{x_g}(\theta)$ and its partial derivative with respect to θ by means of (3) and (4).
4. Select three arbitrary probabilities, p_1, p_2 and p_3, which are in an ascending order, and find out θ_1, θ_2 and θ_3, at which $\hat{M}_{x_g}(\theta)$ equals p_1, p_2 and p_3, respectively.
5. From (12) the estimated acceleration parameter $\hat{\xi}_{x_g}$ is obtained as the solution of

$$\frac{\theta_3 - \theta_2}{\theta_2 - \theta_1} = \frac{\log[(p_2)^{-1/\xi_{x_g}} - 1] - \log[(p_3)^{-1/\xi_{x_g}} - 1]}{\log[(p_1)^{-1/\xi_{x_g}} - 1] - \log[(p_2)^{-1/\xi_{x_g}} - 1]}. \tag{18}$$

6. Obtain the estimate, $\hat{\beta}_{x_g}$, as the solution of

$$\hat{M}_{x_g}(\beta_{x_g}) = \left[\frac{1}{2}\right]^{\xi_{x_g}}. \tag{19}$$

7. From these results obtain the estimate of α_{x_g} by

$$\hat{\alpha}_{x_g} = \frac{2^{\hat{\xi}_{x_g}+1}}{D\hat{\xi}_{x_g}} \frac{\partial}{\partial\theta} \hat{M}_{x_g}(\theta) \qquad \text{at } \theta = \hat{\beta}_{x_g}. \tag{20}$$

Note that this method can be applied for *any* curve as long as $\partial/\partial\theta\, \hat{M}_{x_g}(\theta)$ is available.

Suggestion of the initial nonparametric estimation of the operating characteristics comes from the fact that in the acceleration model the parameters belong to the processing functions, $M_{x_g}(\theta)$'s, rather than the operating characteristics, $P_{x_g}(\theta)$'s, and that the operating characteristics can be estimated more directly and easily from data, and that nonparametric estimation enables us to *discover* the shapes of the actual functions without molding them into some mathematical function as parametric estimation does.

Criteria for Evaluating Models

The general graded response model includes many specific mathematical models. In an effort to select a right model, or models, for a specific psychological reality, the following features will be considered as desirable.

1. The principle behind the model and the set of accompanied assumptions agree with the *psychological reality* in question.

   This indeed is *by far the most important* criterion. The *mathematical modeling* and the *curve fitting* are two different things, although many research papers have been published in which goodnesses of fit of the estimated operating characteristics based on two or more mathematical models to the data are evaluated, and their comparison is used as the *sole* criterion for accepting one of the models and rejecting the other. Even though the fit is good, if the principle behind the model and the set of accompanied assumptions contradict the psychological reality in question, acceptance of such a model is hardly justifiable in scientific research; if it is accepted and research is continued, then we will eventually come across a dead-end street with no meaningful findings. If they agree with the psychological reality, then it should be accepted and research should proceed, unless the fit is intolerably bad. (A good example will be given in a later section, which includes comparison of Figure 8 with Figure 1.)

2. The model provides *additivity* of the operating characteristics of the item scores x_g's.

   Additivity holds if the operating characteristics belong to the same mathematical model under *finer* recategorizations and *combinings* of two or more categories together. This implies that the unique maximum condition, which will be discussed soon, is satisfied by the resulting operating characteristics, if it is satisfied by those of the original x_g's.

   This criterion is all the more important if a model be selected for cognitive assessment with the kind of *rich set* of data described above, for we come across constant *discoveries* of new subprocesses or sets of subprocesses as research proceeds. Even under other circumstances, graded item scores, or partial credits, are more or less *incidental*. For example, on college campuses, it is a general practice to reevaluate the grades, A, B, C, D, and F, in a required course to *pass* and *fail*. Also, with the advancement of computer technologies, it is quite possible to obtain more abundant information from the individual's performance in computerized experiments as we proceed in research, and thus we need finer recategorizations of the whole cognitive process.

3. The model can be *naturally generalized to a continuous response model.*

   This criterion is a natural extension of *additivity*. Examples of such models can be seen in the normal ogive and logistic models in the homogeneous case of the graded response model (Samejima, 1969, 1972), which can be naturally expanded to the normal ogive and logistic models on the continuous response level (Samejima, 1973a), respectively. In general, in the homogeneous case, the *cumulative operating density characteristic*, $P^*_{z_g}(\theta)$, and the *operating density characteristic*, $H_{z_g}(\theta)$, for the continuous response z_g are defined by

$$P^*_{z_g}(\theta) = \int_{-\infty}^{z_g} \psi(u)\, du$$

and

$$H_{z_g}(\theta) = \lim_{\Delta z_g \to 0} \frac{P^*_{z_g}(\theta) - P^*_{(z_g+\Delta z_g)}(\theta)}{\Delta z_g} = a_g \psi(a_g(\theta - b_{z_g}))\left[\frac{db_{z_g}}{dz_g}\right],$$

respectively, where $\psi(\bullet)$ is some density function, and b_{z_g} is the difficulty parameter for the continuous response z_g and is a strictly increasing function of z_g (Samejima, 1973a).

4. The model satisfies the *unique maximum condition* (Samejima, 1969, 1972).

A sufficient, though not necessary, condition that the likelihood function provides a unique maximum for every conceivable response pattern is that

$$\frac{\partial}{\partial \theta} A_{x_g}(\theta) < 0, \tag{21}$$

and

$$\begin{cases} \lim_{\theta \to -\infty} A_{x_g}(\theta) \geq 0 \\ \lim_{\theta \to \infty} A_{x_g}(\theta) \leq 0, \end{cases} \tag{22}$$

for every x_g to every item g. From the condition given by (21), it is obvious that an equality does *not* hold for *both* asymptotes of $A_{x_g}(\theta)$ in (22) for a single x_g. When an equality holds for the same asymptote for all elements of v, a *terminal maximum* will be obtained for the likelihood function, and $\hat{\theta}_v$ assumes one of the two extreme values of θ. It is obvious that (21) can also be expressed by

$$I_{x_g}(\theta) > 0, \tag{23}$$

where $I_{x_g}(\theta)$ is the item response information function given by (6). Thus (21) and (23) can be used interchangeably when the unique maximum condition is discussed. Since $P_{x_g}(\theta)$ is a bounded function between 0 and 1, for $A_{x_g}(\theta)$ to be strictly decreasing in θ, $\partial/\partial\theta\, P_{x_g}(\theta)$ cannot be zero at more than one finite value of θ, and hence in (21) and (23) a strict inequality must hold, indicating that it must be either (a) strictly increasing in θ, (b) strictly decreasing in θ, or (c) unimodal, and at no points its partial derivative equals zero except for the modal point in (c). The fact that $P_{x_g}(\theta)$ is a bounded function further implies that

$$\lim_{\theta \to -\infty} \frac{\partial}{\partial \theta} P_{x_g}(\theta) = \lim_{\theta \to \infty} \frac{\partial}{\partial \theta} P_{x_g}(\theta) = 0,$$

indicating that, in (22), an equality must hold in the second formula for (a), in the first formula for (b), and in neither for (c), and

$$\begin{cases} \lim_{\theta \to -\infty} P_{x_g}(\theta) = 0 & \text{for (a) and (c)} \\ \lim_{\theta \to \infty} P_{x_g}(\theta) = 0 & \text{for (b) and (c).} \end{cases}$$

For simplicity, this set of conditions is called the *unique maximum condition* (Samejima, 1972), and these three types of $P_{x_g}(\theta)$'s which satisfy the unique maximum condition are said to be of types i, ii and iii, respectively. Satisfaction of this condition assures that the likelihood function, given by (1), of *any* response pattern consisting of such response categories has a unique local or terminal maximum.

Not only satisfaction of the unique maximum condition assures the uniqueness of the maximum likelihood estimate, but also it contributes to the identification of the uniqueness of other estimates, which maximizes

$$w(\theta)L(v|\theta),$$

where $w(\theta)$ is some weight function, for the likelihood equation can be written as

$$\frac{\partial}{\partial\theta}\log L(v|\theta) + \frac{\partial}{\partial\theta}\log w(\theta) = 0.$$

Thus, if the unique maximum condition is satisfied, then the additional condition is that $\partial/\partial\theta \log w(\theta)$ satisfies the analogous condition as the one assigned for $A_{x_g}(\theta)$. These estimators include the Bayes modal estimator (Samejima, 1969), where a prior is used for $w(\theta)$, and Warm's weighted likelihood estimator (Warm, 1989), whose $w(\theta)$ does not depend on any specific ability distribution.

5. The model provides the *ordered modal points* of the operating characteristics in accordance with the item scores.

   Using the basic function defined by (5), a sufficient, though not necessary, condition for the strict orderliness of the modal points of the operating characteristics, $P_{x_g}(\theta)$'s, is that

$$A_{(x_g-1)}(\theta) < A_{x_g}(\theta) \tag{24}$$

   for all θ for $x_g = 1, 2, \ldots, m_g$.

From its definition it is obvious that for any mathematical model in the *homogeneous case*:

1. *Additivity* of the operating characteristics *always holds*.
2. A natural expansion of the model to a *continuous* response model can be done.
3. If the unique maximum condition is satisfied, then a strict orderliness among the modal points of $P_{x_g}(\theta)$'s also holds, for it can be shown (Samejima, 1972) that the relationship,

$$A_{(x_g-1)}(\theta) < \tilde{A}_{x_g}(\theta) < A_{x_g}(\theta),$$

   holds for $x_g = 1, 2, \ldots, m_g$, where $\tilde{A}_{x_g}(\theta)$ is the *asymptotic basic function* (Samejima, 1969, 1972) defined in the homogeneous case by

$$\tilde{A}_{x_g}(\theta) \equiv \lim_{\lambda_{(x_g+1)}\to\lambda_{x_g}} A_{x_g}(\theta) = \frac{\partial}{\partial\theta}\log\left[\frac{\partial}{\partial\theta}P^*_{x_g}(\theta)\right] = \frac{\dfrac{\partial^2}{\partial\theta^2}M_1(\theta-\lambda_{x_g})}{\dfrac{\partial}{\partial\theta}M_1(\theta-\lambda_{x_g})},$$

with λ_{x_g} being zero for $x_g = 1$ and increases with x_g, which is identical in shape for all $x_g = 1, 2, \ldots, m_g$ except for the position alongside the dimension θ. Thus (24) is satisfied.

It has been demonstrated (Samejima, 1969, 1972) that both the normal ogive model and the logistic model belong to this class of models.

While models in the homogeneous case that satisfy the unique maximum condition have most of the other desirable features also, those in the *heterogeneous* case fulfillment of these criteria becomes more difficult. Models in the heterogeneous case tend to provide greater varieties in the configuration of the operating characteristics, however. This implies that search of a model in the heterogeneous case may be more successful in obtaining one which satisfies the most important criterion, that is, agreement of the rationale behind the model with the psychological reality in question.

Soundness and Robustness of the Acceleration Model

If we observe the acceleration model in terms of each of the five criteria, the following will be noted.

1. The principle behind the model and the set of accompanied assumptions fit problem solving and other cognitive processes fairly well, as well as many other psychological processes that involve ordered categories or partial credits.
2. The model *can be generalized to a continuous response model* as the limiting situation in which there are infinitely many subprocesses within each step.

As for each of the other three criteria, observations and discussion will be made for the special case in which (12) is used for the processing function.

Additivity

If our experimental setting is improved and allows us to observe the individual's performance in more finely graded steps, then m_g will become larger. It is obvious from (12) and the definition of ξ_{x_g} that the resulting operating characteristics still belong to the acceleration model: a partial satisfaction of the *additivity* criterion.

Suppose, on the contrary, we need to combine two steps which do not belong to the same cluster. Note that the resulting combined step will *not* belong to the acceleration model. By virtue of the *robustness* of the acceleration model, however, in most cases the operating characteristic of the combined step can be well approximated by the acceleration model. Thus *additivity* of the operating characteristics *practically* holds in this model.

To illustrate this, Figure 2 presents six step processing functions, the first three of which belong to a cluster with $\alpha_{x_g} = 1.0$ and $\beta_{x_g} = -1.0$, and the second three to another with $\alpha_{x_g} = 1.0$ and $\beta_{x_g} = 1.0$, respectively, and the third parameters are $\xi_{x_g} = 0.5, 1.0, 1.5$, for the three steps in each cluster.

It is obvious that the operating characteristic of the combined category of any two adjacent steps still belongs to this specific acceleration model, except for the combination of $x_g = 2$ and $x_g = 3$, for it does not belong to (12). This product (solid line) is also shown in Figure 2, and so is its approximation (a dash and two dots repeated) obtained by fitting a single $M_{x_g}(\theta)$ through (18), (19) and (20). Figure 3 presents the operating characteristics of the seven steps, 0 through 6, plus the *sum* of the two operating characteristics for $x_g = 2$ and $x_g = 3$ (solid line) and the approximated operating characteristic (a dash and two dots repeated) obtained by using the approximated $M_{x_g}(\theta)$, which was fitted to the product of $M_3(\theta)$ and $M_4(\theta)$. The values of p_1, p_2, p_3

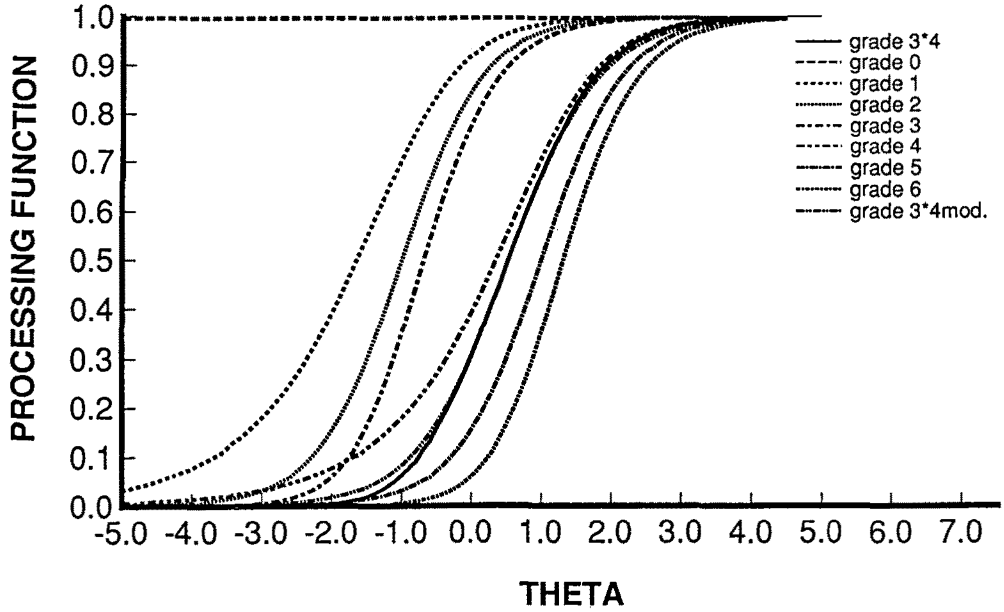

FIGURE 2.
Six step processing functions, three of which belong to one cluster and the other three to another cluster, plus the product of the two processing functions for $x_g = 3$ and $x_g = 4$ (solid line) and its approximation by the processing function in the acceleration model (a dash and two dots repeated).

used in this approximation were 0.21109, 0.48884, 0.79446, and the corresponding θ_1, θ_2, θ_3 were −0.3, 0.5, 1.4. The resulting estimated parameters turned out to be: $\hat{\xi}_{x_g}$ = 1.11338, $\hat{\beta}_{x_g}$ = 0.43006, and $\hat{\alpha}_{x_g}$ = 0.86888. The two curves for the combined category in Figure 3 overlap almost completely, showing the *robustness* of the model.

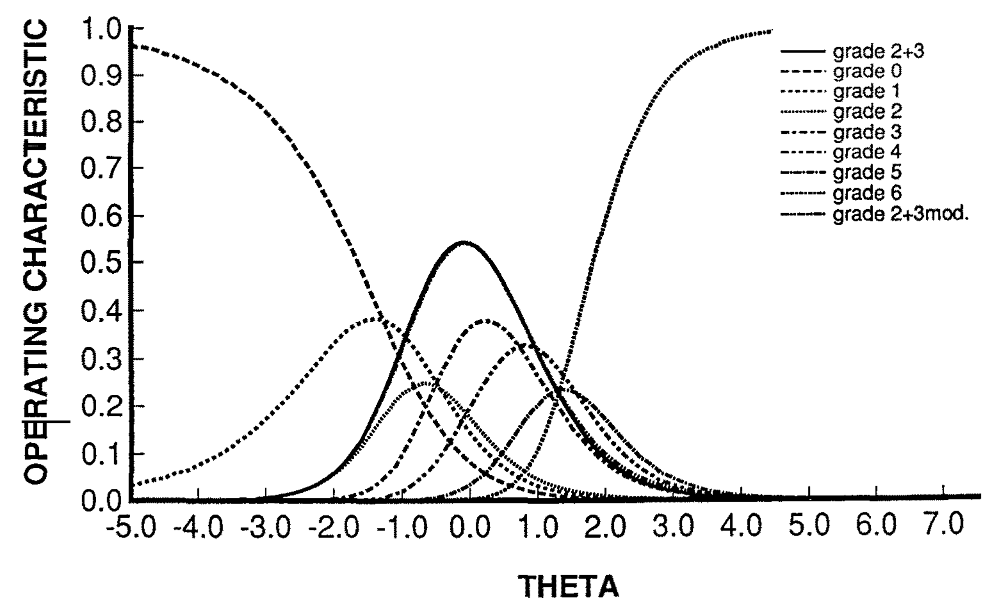

FIGURE 3.
Operating characteristics of seven steps plus the sum of two operating characteristics for $x_g = 2$ and $x_g = 3$ (solid line) and the approximated operating characteristic for this combined category in the acceleration model (a dash and two dots repeated).

The reason for the robustness of this model comes from the fact that the two parameters, α_{x_g} and ξ_{x_g}, work *compensatorily* to determine the *steepness* of $M_{x_g}(\theta)$, while ξ_{x_g} *alone* accounts for the shape of the curve. Thus a set of a large α_{x_g} and a small ξ_{x_g} will provide the steepness of the curve similar to the one resulting from a set of a small α_{x_g} and a large ξ_{x_g}. The shape of the curve is largely determined by ξ_{x_g}, however, as we can see in the earlier observation that θ_{dmax} changes as a function of ξ_{x_g}, thus together providing various shapes and steepnesses.

Robustness of the model also saves the situation in which the assumption that a single set of α_{x_g} and β_{x_g} exists within each step is violated, which may occur especially when m_g is small. Suppose that we did *not* know there were *two* clusters involved, and treated them as a single step, estimating the step parameters following this specific acceleration model, and, later, with the improvement of the experimental setting, they were disclosed as two separate steps which belonged to two different clusters. The result obtained by treating them as a single step still provides *good approximations*, as illustrated in Figure 3.

Satisfaction of the Unique Maximum Condition

From (15), the two asymptotes of the basic function are given by

$$\begin{cases} \lim_{\theta\to-\infty} A_{x_g}(\theta) = D \sum_{u\le x_g} \xi_u \alpha_u > 0 & \text{for } x_g = 1, 2, \ldots, m_g \\ \lim_{\theta\to\infty} A_{x_g}(\theta) = -D\alpha_{x_g+1} < 0 & \text{for } x_g = 0, 1, \ldots, m_g - 1, \end{cases} \tag{25}$$

and both the upper asymptote for $x_g = 0$ and the lower asymptote for $x_g = m_g$ become zero. Although the first line of (25) can be obtained easily, the second line needs some work. Since ξ_{x_g} is a positive real number, we can find a rational number within the interval, $(\xi_{x_g} - \varepsilon, \xi_{x_g})$, and also within the interval, $(\xi_{x_g}, \xi_{x_g} + \varepsilon)$, however small $\varepsilon(>0)$ may be. Thus ξ_{x_g} is considered as the limit of these two rational numbers when ε tends to zero.

Let k and m be positive integers, which satisfy

$$\left|\frac{k}{m} - \xi_{x_g+1}\right| < \varepsilon, \tag{26}$$

and define

$$\Gamma_{x_g}(\theta) = [\Psi_{x_g}(\theta)]^{1/m}. \tag{27}$$

From (15) and (27) we obtain

$$\lim_{\theta\to\infty} A_{x_g}(\theta) = -D\alpha_{x_g+1} \lim_{\varepsilon\to0} \lim_{\theta\to\infty} \left[\frac{\frac{k}{m}[\Gamma_{(x_g+1)}(\theta)]^k \sum_{u=0}^{m-1} [\Gamma_{(x_g+1)}(\theta)]^u}{\sum_{u=0}^{k-1} [\Gamma_{(x_g+1)}(\theta)]^u}\right]$$

$$= -D\alpha_{x_g+1},$$

the result shown in the second line of (25).

To prove that (21) holds with this model, from (5) it is sufficient to show that

$$\begin{cases} \dfrac{\partial^2}{\partial \theta^2} \log M_{x_g}(\theta) < 0 \\ \dfrac{\partial^2}{\partial \theta^2} \log [1 - M_{x_g}(\theta)] < 0, \end{cases} \tag{28}$$

for $x_g = 1, 2, \ldots, m_g$. From (12), (13) and (14) we can write

$$\frac{\partial^2}{\partial \theta^2} \log M_{x_g}(\theta) = -\xi_{x_g} D^2 \alpha_{x_g}^2 \Psi_{x_g}(\theta)[1 - \Psi_{x_g}(\theta)]$$

$$= -D^2 \sum_{w \in x_g} \xi_w \alpha_{x_g}^2 \Psi_{x_g}(\theta)[1 - \Psi_{x_g}(\theta)] < 0,$$

for $x_g = 1, 2, \ldots, m_g$, that is, the first line of (28) has been demonstrated.

To prove the second line of (28), it is sufficient to show

$$\frac{\partial}{\partial \theta} \log \left[\frac{\partial}{\partial \theta} M_{x_g}(\theta)\right] - \frac{\partial}{\partial \theta} \log [1 - M_{x_g}(\theta)] > 0. \tag{29}$$

From (12), (13), (14) and (26) with the replacement of ξ_{x_g+1} by ξ_{x_g}, we have

$$\frac{\partial}{\partial \theta} \log \left[\frac{\partial}{\partial \theta} M_{x_g}(\theta)\right] = D\alpha_{x_g}[\xi_{x_g}(1 - \Psi_{x_g}(\theta)) - \Psi_{x_g}(\theta)],$$

$$\frac{\partial}{\partial \theta} \log [1 - M_{x_g}(\theta)] = -\frac{\xi_{x_g} D\alpha_{x_g}[\Psi_{x_g}(\theta)]^{\xi_{x_g}}[1 - \Psi_{x_g}(\theta)]}{1 - [\Psi_{x_g}(\theta)]^{\xi_{x_g}}};$$

and then

$$\frac{\partial}{\partial \theta} \log \left[\frac{\partial}{\partial \theta} M_{x_g}(\theta)\right] - \frac{\partial}{\partial \theta} \log [1 - M_{x_g}(\theta)]$$

$$= \frac{D\alpha_{x_g}[\xi_{x_g}\{1 - \Psi_{x_g}(\theta)\} - \Psi_{x_g}(\theta)\{1 - [\Psi_{x_g}(\theta)]^{\xi_{x_g}}\}]}{1 - [\Psi_{x_g}(\theta)]^{\xi_{x_g}}}$$

$$= \lim_{\varepsilon \to 0} \frac{D\alpha_{x_g}}{\sum_{s=0}^{k-1} [\Gamma_{x_g}(\theta)]^s} \sum_{t=0}^{k-1} \left[\frac{1}{m} \sum_{u=0}^{m-1} [\Gamma_{x_g}(\theta)]^u - [\Gamma_{x_g}(\theta)]^{m+t}\right] > 0,$$

that is, satisfaction of (29), and hence of the second line of (28). Thus (21) holds throughout the entire range of θ for $x_g = 0, 1, 2, \ldots, m_g$. *The unique maximum condition is satisfied*, therefore, for this specific acceleration model.

Orderliness of the Modal Points of the Operating Characteristics

The orderliness of the modal points of the operating characteristics *practically* holds in this model, except for certain unusual cases. In general, we have

$$A_{(x_g+1)}(\theta) - A_{x_g}(\theta) = \frac{\frac{\partial}{\partial\theta} M_{x_g}(\theta)}{M_{x_g}(\theta)} - \frac{-\frac{\partial}{\partial\theta} M_{x_g}(\theta)}{1 - M_{x_g}(\theta)} + \frac{-\frac{\partial}{\partial\theta} M_{(x_g+1)}(\theta)}{1 - M_{(x_g+1)}(\theta)} \tag{30}$$

$$= \frac{\frac{\partial}{\partial\theta} M_{x_g}(\theta)}{M_{x_g}(\theta)(1 - M_{x_g}(\theta))} - \frac{\frac{\partial}{\partial\theta} M_{(x_g+1)}(\theta)}{1 - M_{(x_g+1)}(\theta)};$$

and from (12), (13) and (27) we can write

$$\frac{\frac{\partial}{\partial\theta} M_{x_g}(\theta)}{M_{x_g}(\theta)(1 - M_{x_g}(\theta))} = D\alpha_{x_g} \lim_{\varepsilon\to 0} \frac{\frac{1}{m}\sum_{t=0}^{m-1} [\Gamma_{x_g}(\theta)]^t}{\frac{1}{k}\sum_{u=0}^{k-1} [\Gamma_{x_g}(\theta)]^u}, \tag{31}$$

and

$$-\frac{\frac{\partial}{\partial\theta} M_{(x_g+1)}(\theta)}{1 - M_{(x_g+1)}(\theta)} = -D\alpha_{x_g+1} \lim_{\varepsilon\to 0} [\Gamma_{(x_g+1)}(\theta)]^h \frac{\frac{1}{r}\sum_{t=0}^{r-1} [\Gamma_{x_g+1}(\theta)]^t}{\frac{1}{h}\sum_{u=0}^{h-1} [\Gamma_{x_g+1}(\theta)]^u}. \tag{32}$$

Substituting (31) and (32) into (30), we obtain

$$A_{(x_g+1)}(\theta) - A_{x_g}(\theta) = D\left[\alpha_{x_g} \lim_{\varepsilon\to 0} \frac{\frac{1}{m}\sum_{t=0}^{m-1} [\Gamma_{x_g}(\theta)]^t}{\frac{1}{k}\sum_{u=0}^{k-1} [\Gamma_{x_g}(\theta)]^u} - \alpha_{x_g+1} \lim_{\varepsilon\to 0} [\Gamma_{(x_g+1)}(\theta)]^h \frac{\frac{1}{r}\sum_{t=0}^{r-1} [\Gamma_{x_g+1}(\theta)]^t}{\frac{1}{h}\sum_{u=0}^{h-1} [\Gamma_{x_g+1}(\theta)]^u}\right]. \tag{33}$$

Thus possibilities of nonorderliness exist if (33) is less than zero at the value of θ where $A_{x_g}(\theta) = 0$. This can happen, for example, when $\xi_{x_g+1} \ll \xi_{x_g}$ and $\beta_{x_g} \ll \beta_{x_g+1}$. Note, however, that in such a situation $M_{(x_g+1)}(\theta)$ becomes very flat, and the unidimensionality of this *step* should be questioned. This is rather an unusual case, and, in practice, it is expected that *orderliness of the modal points* of the operating characteristics *usually* holds.

Figures 4 through 7 illustrate the case in which reversal of the modal points occurs between two adjacent item scores, showing the *processing functions*, the *cumulative*

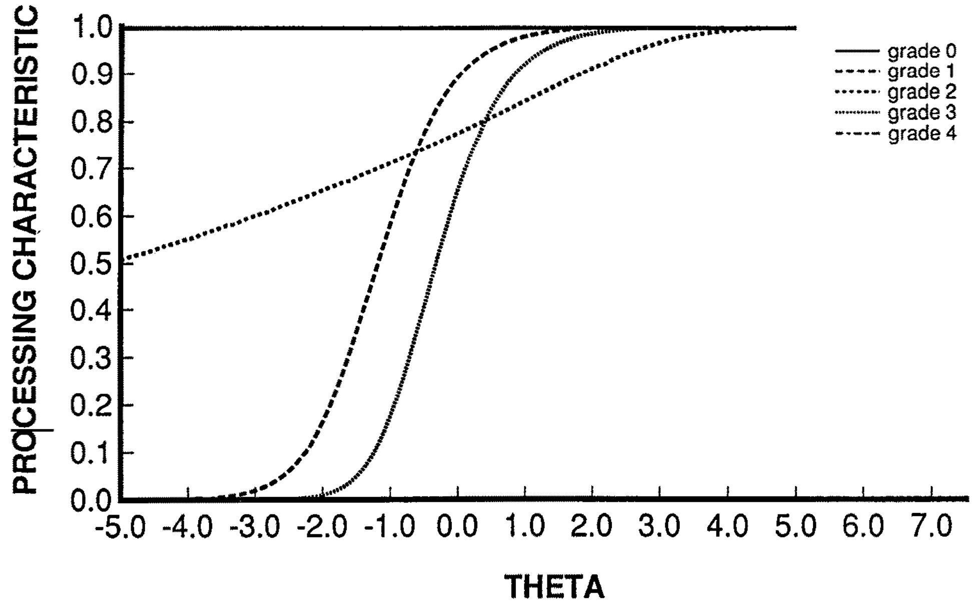

FIGURE 4.
Example of a set of processing functions of five steps in the acceleration model. The processing function for $x_g = 2$ is unusually flat.

operating characteristics, the *operating characteristics* and the basic functions. In this example, $m_g = 3$, $\alpha_{x_g} = 1.0$ for all x_g's, and $\beta_1 = -1.5$, $\beta_2 = 3.0$ and $\beta_3 = -1.0$, and the acceleration parameters are 1.5, 0.05, 2.5 for $x_g = 1, 2, 3$, respectively. From Figure 7 it is obvious that the relationship between the values of $A_1(\theta)$ and $A_2(\theta)$ is

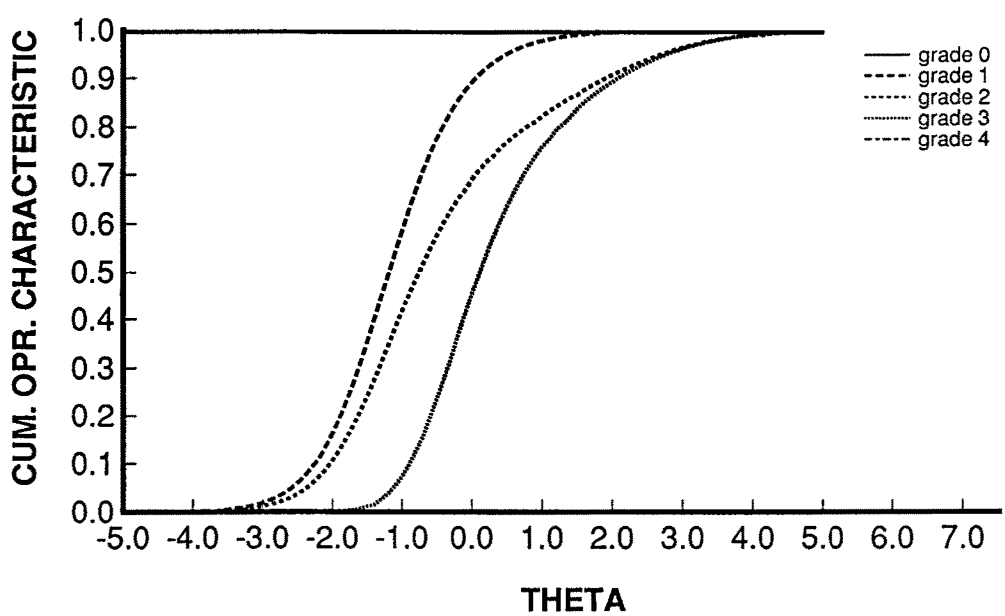

FIGURE 5.
The set of five cumulative operating characteristics in the acceleration model, resulting from the processing functions illustrated in Figure 4.

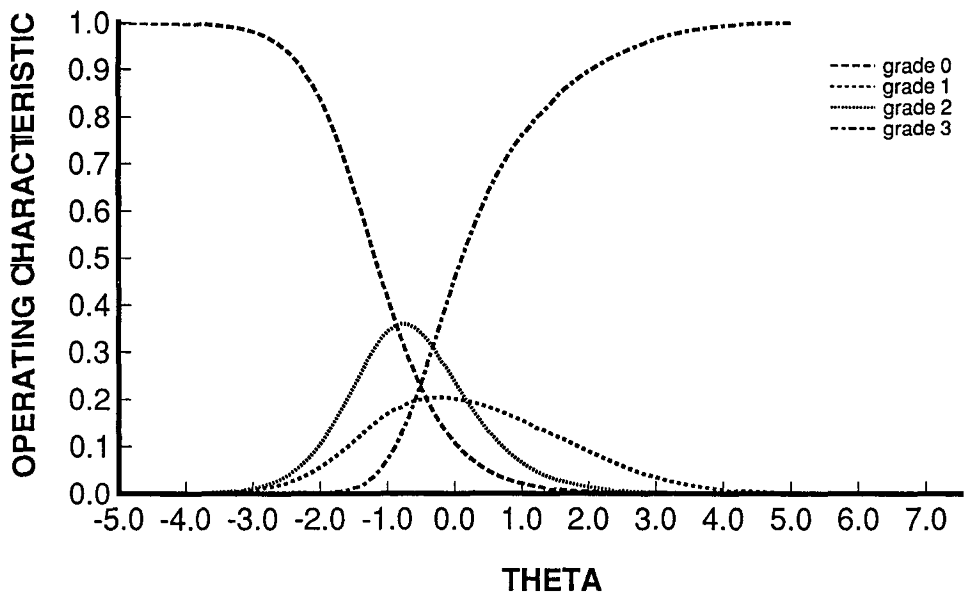

FIGURE 6.
The set of four operating characteristics in the acceleration model, resulting from the processing functions illustrated in Figure 4. The modal points for $x_g = 1$ and $x_g = 2$ are reversed.

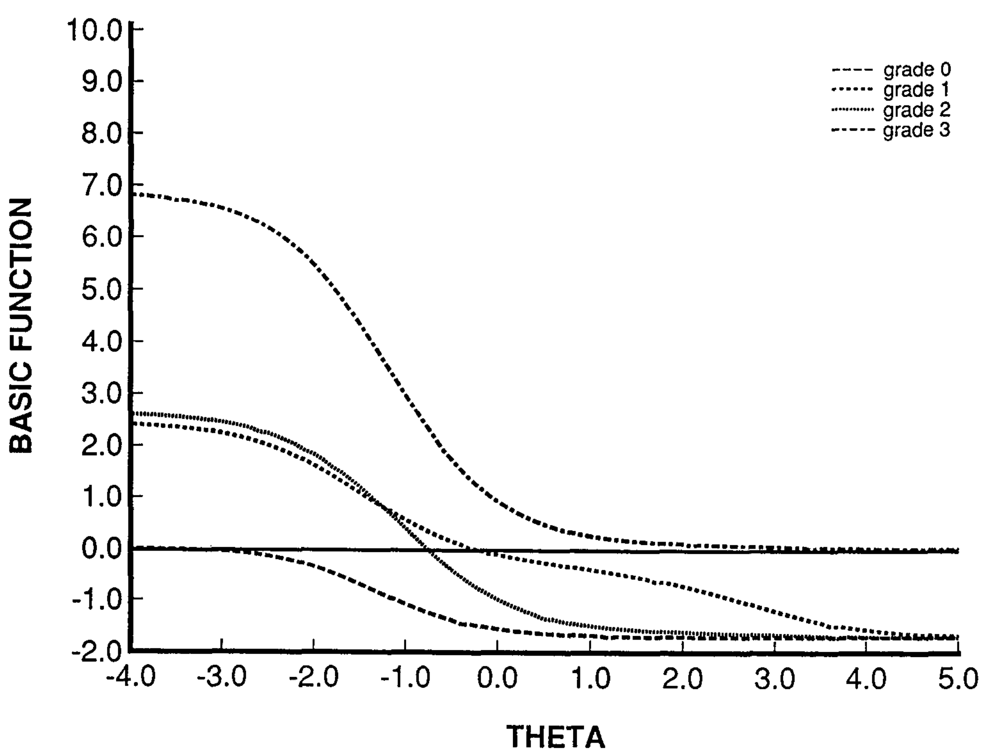

FIGURE 7.
The set of four basic functions in the acceleration model, resulting from the processing functions illustrated in Figure 4. The two curves for $x_g = 1$ and $x_g = 2$ are reversed before reaching the abscissa, causing the reversal of the modal points of the operating characteristics.

reversed before either of them equals 0 and thus the reversal of the two modal points in Figure 6 occurs.

Graded Response Models Based on Individual Choice Behavior

Samejima (1972) has pointed out that Bock's multinomial model (Bock, 1972), represented by

$$P_{k_g}(\theta) = \frac{\exp\,[\alpha_{k_g}\theta + \beta_{k_g}]}{\sum_{u \in K_g} \exp\,[\alpha_u \theta + \beta_u]}, \tag{34}$$

where k_g denotes a nominal response to item g and $\alpha_{k_g}(>0)$ and β_{k_g} are item response parameters, can be considered as a graded response model in the heterogeneous case, if the nominal response k_g in (34) is replaced by the graded item response x_g and the parameter α_{x_g} satisfies

$$\alpha_0 \le \alpha_1 \le \alpha_2 \le \cdots \le \alpha_{m_g}, \tag{35}$$

where a strict inequality should hold at least at one place. The basic function $A_{x_g}(\theta)$ in this model is obtained from (5) and (34) by

$$A_{x_g}(\theta) = \frac{\sum_{u=0}^{m_g} [\alpha_{x_g} - \alpha_u] \exp\,[\alpha_u \theta + \beta_u]}{\sum_{u=0}^{m_g} \exp\,[\alpha_u \theta + \beta_u]} = \alpha_{x_g} - \frac{\sum_{u=0}^{m_g} \alpha_u \exp\,[\alpha_u \theta + \beta_u]}{\sum_{u=0}^{m_g} \exp\,[\alpha_u \theta + \beta_u]};$$

and the item response information function can be written from (6) and (34) as

$$I_{x_g}(\theta) = \frac{\sum_{s=0}^{m_g} \alpha_s^2 \exp\,[\alpha_s \theta + \beta_s]}{\sum_{u=0}^{m_g} \exp\,[\alpha_u \theta + \beta_u]} + \left[\frac{\sum_{s=0}^{m_g} \alpha_s \exp\,[\alpha_s \theta + \beta_s]}{\sum_{u=0}^{m_g} \exp\,[\alpha_u \theta + \beta_u]}\right]^2 = I_g(\theta), \tag{36}$$

which is identical for every x_g, 0 through m_g, and thus equals the item information function $I_g(\theta)$.

It has been demonstrated (Samejima, 1972) that the model satisfies the *unique maximum condition*, and the perfect *orderliness of the modal points* of the operating characteristics is realized if in (35) strict inequality holds between *every* pair of α_{x_g}'s, or, otherwise, the same modal point is shared by two or more x_g's whose α_{x_g}'s are equal, indicating that multi-incorrect responses exist when such x_g's include 0, and multi-correct responses exist when they include m_g.

Samejima did not pursue those models, however, for the reason that Bock's model is based upon the assumption that the conditional ratio, given θ, of the probabilities of any two discrete responses to item g is *invariant* regardless of the set of alternatives selected from the answer space, the same assumption used in the individual choice behavior (Luce, 1959). This ratio is given by

$$\frac{P_{k_g}(\theta)}{P_{h_g}(\theta)} = \frac{\exp[\alpha_{k_g}\theta + \beta_{k_g}]}{\exp[\alpha_{h_g}\theta + \beta_{h_g}]} = \exp[(\alpha_{k_g} - \alpha_{h_g})\theta] \exp[\beta_{k_g} - \beta_{h_g}], \quad (37)$$

where h_g is a discrete response to item g which is different from k_g. If we translate to the graded response model, this assumption requires that, if we add B+ between A and B in the original grading system of A, B, C, D and F, for example, the ratio of the conditional probabilities for, say, B and F, given θ, be unchanged: the assumption which *does not fit* the reality.

Bock used his nominal model for multiple-choice test items and discovered their *implicit* orders among the *distractors*. The assumption is legitimate in this application. Let us consider a *distractor space*, which is a subspace of the answer space and elements of which can be used as distractors of a specific multiple-choice question. We can select a subset of the distractor space as the actual distractors for the item, or some other subset. Suppose that the two subsets have two distractors, k_g and h_g, in common, and all the others are different. If we consider answering the multiple-choice test item as a choice behavior, it is reasonable to assume that the operating characteristics, $P_{k_g}(\theta)$ and $P_{h_g}(\theta)$, are different when put in different subsets, but the conditional ratio, given θ, of selecting k_g and h_g, which is given by (37), stays the same. Thus the item response parameters for k_g and h_g are meaningful, and can be used across two or more subsets of distractors which include k_g and h_g. After the item response parameters have been estimated for every member of the distractor space, we shall be able to use any subset of the distractor space as graded responses, ordering them in terms of the parameters α_{k_g}'s, as shown in (35).

This assumption is *not* acceptable in typical graded response or partial credit situations, however, as was exemplified above. Thus applicabilities of those models based upon the individual choice behavior are very narrowly limited, for *additivity of the operating characteristics and the generalizability to a continuous response model do not hold* in such a graded response model.

Also the meanings of item parameters in those models become unclear. For example, Masters (1982) has proposed a partial credit model whose operating characteristic of x_g is a special case of (34) in which

$$\alpha_{x_g} = x_g + 1$$

for $x_g = 0, 1, 2, \ldots, m_g$, and δ_{x_g} is defined to satisfy

$$\beta_{x_g} = -\sum_{u=0}^{x_g} \delta_u .$$

Muraki (1992) has proposed the generalized partial credit model, the operating characteristic of which is another special case of (34), and in which a_g and b_{x_g} are defined to satisfy

$$\alpha_{x_g} = (x_g + 1)a_g$$

and

$$\beta_{x_g} = -a_g \sum_{u=0}^{x_g} b_u .$$

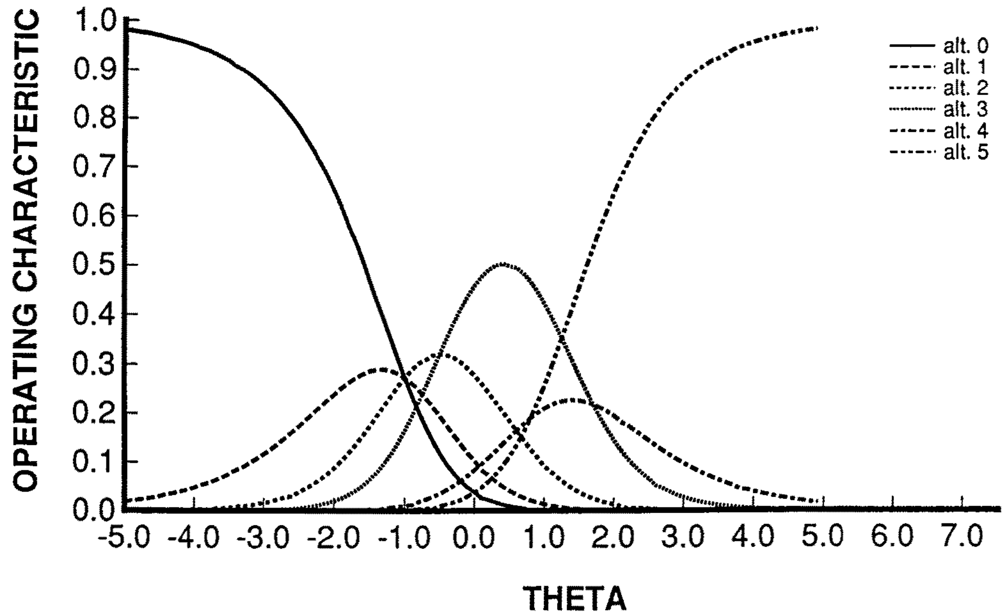

FIGURE 8.
Example of a set of operating characteristics of six item scores in Masters' partial credit model.

A problem common in these two models lies in the fact that α_{x_g} is defined as a function of the graded score x_g, which is *incidental*. Thus a change in the grading system easily affects the value of this parameter, and the meaning of the item response parameter itself becomes unclear.

Suppose, for example, a model based on the individual choice behavior, or any other model, has been used, and then the researcher decides to *switch* to the specific acceleration model for further research. In such a case, the method represented by (18), (19) and (20) can be used directly. Figure 8 presents the operating characteristics in Masters' partial credit model, using $\alpha_{x_g} = 1, 2, 3, 4, 5, 6$ and $\beta_{x_g} = 1.0, 2.0, 3.0, 3.5, 1.8, 1.0$ in (34) with k_g replaced by x_g for $x_g = 0, 1, 2, 3, 4, 5$, respectively.

In fact, the parameters in the acceleration model used in Figure 1 were obtained by (18), (19) and (20) from the $M_{x_g}(\theta)$'s in Masters' model with the above set of parameters, setting $p_1 = 0.1$, $p_2 = 0.5$ and $p_3 = 0.9$ in (18). Compare Figure 8 with Figure 1. These two sets of curves are practically indistinguishable! (For this reason, in Figure 8 *alt.* is used instead of *grade* to stand for an *alternative* in the multiple-choice test item following Bock's application, to make these figures distinguishable.) The same procedure was repeated by setting the values of p_1, p_2 and p_3 to (0.2, 0.5, 0.8), (0.25, 0.50, 0.75), (0.3, 0.5, 0.7) and (0.3, 0.6, 0.7), respectively, and the results turned out to be very similar. The item information function in the acceleration model (solid line) with the parameter values used for Figure 1, and that in Masters' partial credit model (dashed line) obtained by substituting the above parameter values into (36), are shown in Figure 9. These two curves are reasonably close to each other.

Discussion

General graded response model was discussed, and the acceleration model was introduced as a mathematical model in the heterogeneous case which is sound and useful in cognitive assessment as well as in analyzing more traditional psychometric data. The robustness of the model was also discussed.

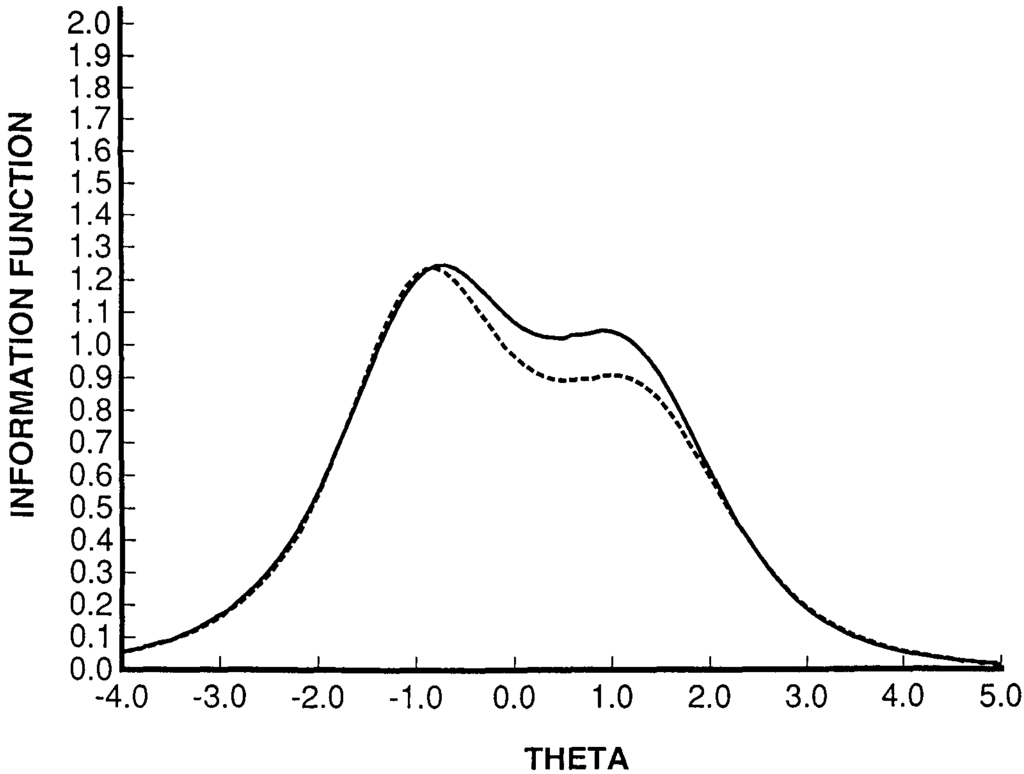

FIGURE 9.
Comparison of the item information functions in Masters' model (dashed line) and in the acceleration model (solid line).

If the majority of the processing functions are flat for a certain interval of θ, then a strictly increasing scale transformation may solve the problem. If a very flat $M_{x_g}(\theta)$ is obtained for one or more x_g's, it should be taken as an indicator of possible violation of the unidimensionality. It is possible that some of the subprocesses require ability other than θ, and, if this is the case, then the assumption of local independence in the unidimensional latent space will be violated; it is necessary to turn to the model expanded to a multidimensional latent space. Expansion of the acceleration model can be done in a similar manner to the way in which the normal ogive model was expanded (Samejima, 1974). It will be wise to wait, however, until the unidimensional model has been used with empirical data, and then we come across situations in which the expansion is necessary, before the decision is made as to in what way the model should be expanded to the multidimensional latent space.

Use of some other function, such as the normal ogive function, for $\Psi_{x_g}(\theta)$ may provide substantially different results. This must be examined in the future.

DiBello, Stout and Roussos (1993) have proposed a unified cognitive/psychometric diagnosis, in an effort to bridge psychometric methodologies with cognitive assessment. Samejima (1995) has proposed the competency space approach, using latent trait models in cognitive assessment. In this latter approach, the acceleration model is expected to find its usefulness.

References

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* (chapters 17–20). 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, 37*, 29–51.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1993). *Unified cognitive/psychometric diagnosis foundations and application*. Paper presented at the 1993 American Educational Research Association Meeting, Atlanta, GA.

Levine, M. (1984). *An introduction to multilinear formula scoring theory* (Office of Naval Research Report, 84-4). Washington, DC: Department of the Navy.

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

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

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

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

Ramsay, J. O., & Wang, X. (1993). *Hybrid IRT models*. Paper presented at the 1993 Psychometric Society meeting, Berkeley, California.

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores, *Psychometrika Monograph No. 17, 34*(4, Pt. 2).

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph No. 18, 37*(4, Pt. 2).

Samejima, F. (1973a). Homogeneous case of the continuous response model. *Psychometrika, 38*, 203–219.

Samejima, F. (1973b). A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika, 38*, 221–233.

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

Samejima, F. (1983). Some methods and approaches of estimating the operating characteristics of discrete item responses. In H. Wainer & S. Messick (Eds.), *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord* (pp. 159–182). Hillsdale, NJ: Lawrence Erlbaum.

Samejima, F. (1993). Human psychological behavior: Approached by latent trait models. *Proceedings of the IEEE International Workshop on Neuro-Fuzzy Control, Muroran, Japan*, 26–33.

Samejima, F. (1994). Roles of Fisher type information in latent trait models. In H. Bozdogan (Ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*. Netherlands: Kluwer Academic Publishers. (The Convention was held in May 1992 in Knoxville, TN)

Samejima, F. (1995). A cognitive diagnosis method using latent trait models: competency space approach and its relationship with DiBello and Stout's unified cognitive/psychometric diagnosis model. In P. D. Nichols, S. E. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 391–410). Hillsdale, NJ: Lawrence Erlbaum.

Samejima, F. (in press). Rationale and actual procedures of efficient nonparametric approaches for estimating the operating characteristics of discrete item responses, *Psychometrika*.

Manuscript received 3/29/94
Final version received 9/12/94