

Distinctive and Incompatible Properties of Two Common Classes of IRT Models for Graded Responses

David Andrich

Murdoch University

Two classes of models for graded responses, the first based on the work of Thurstone and the second based on the work of Rasch, are juxtaposed and shown to satisfy important, but mutually incompatible, criteria and to reflect different response processes. Specifically, in the Thurstone models if adjacent categories are joined to form a new category, either before or after the data are collected, then the probability of a response in the new category is the sum of the probabilities of the responses in the original categories. However, the model does not have the explicit property that if the categories are so joined, then the estimate of the location of the entity or object being measured is invariant before and after the joining. For the Rasch models, if a pair of adjacent categories are joined and then the data are collected, the es-

timate of the location of the entity is the same before and after the joining, but the probability of a response in the new category is not the sum of the probabilities of the responses in the original categories. Furthermore, if data satisfy the model and the categories are joined after the data are collected, then they no longer satisfy the same Rasch model with the smaller number of categories. These differences imply that the choice between these two classes of models for graded responses is not simply a matter of preference; they also permit a better understanding of the choice of models for graded response data as a function of the underlying processes they are intended to represent. *Index terms: graded responses, joining assumption, polytomous IRT models, Rasch model, Thurstone model.*

The collection of data in the form of graded responses is ubiquitous in the social sciences (Dawes, 1972); it is used when measurement of the kind used in the physical sciences would be desirable, but no such relevant measuring instrument exists. The prototype of measurement is the partitioning of a continuum, abstracted as the real number line, by equidistant thresholds sufficiently fine that their own width can be ignored, and the location of an entity or object on this line with the use of an instrument. The location of the entity or object, typically called the measurement, is the number of thresholds exceeded by the object from the origin, and is defined in the unit of the common distance between the thresholds. Although measuring instruments have operating ranges, a measurement is not taken to be a function of the operating range of any instrument—instead, if a measurement is contaminated by the operating range of the instrument (e.g., floor or ceiling effects), another instrument with a range more compatible with the location of the entity or object is sought. In addition, all measurements involve error. However, in deterministic frameworks, it is assumed that the error is sufficiently small relative to the measurements that the error can be ignored. This paper draws on this prototype of measurement in analyzing models for graded responses.

Measurement in the Social Sciences

Graded response formats in the social sciences mirror measurement in the physical sciences in that the entity—the person being measured on a given variable—is postulated to be located on a latent continuum

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 19, No. 1, March 1995, pp. 101–119

© Copyright 1995 Applied Psychological Measurement Inc.

0146-6216/95/010101-19\$2.20

101

that is partitioned into contiguous intervals or categories. In elementary treatments, the prototype of measurement described above is followed closely in that the location of the person is taken as the number of categories from the first in a selected direction. For example, traditional analyses of Likert-style questionnaires (Likert, 1932) begin this way. In more advanced treatments, three modifications to this prototype are made: (1) rather than ignoring any errors of classification into categories, an explicit probabilistic response process that is governed by the location of the person and the number and location of thresholds is postulated; (2) therefore, the finiteness of the number of categories is taken into account explicitly; and (3) the distances between thresholds are not assumed to be equal, but are estimated from the data.

Thurstone versus Rasch models. It generally is considered desirable that the formulation of a probabilistic model should reflect the process by which a person is classified into one of the graded responses (McCullagh, 1980, 1985), and a rationale is generally provided in building a model. Two generic classes of models have become dominant in the analysis of typical formats of graded responses. One class is based on the work of Thurstone (Edwards & Thurstone, 1952) and elaborated by Bock (1975), McCullagh (1980, 1985), Samejima (1969), and others; the second is based on the work of Rasch (1961) and elaborated by Andersen (1977), Andrich (1978, 1985), Muraki (1992), Wright and Masters (1982), and others.

These two classes of models have been compared descriptively in terms of the structure of their parameters (e.g., Thissen & Steinberg, 1986) and in terms of the results they provide when applied to the same data (De Ayala, Dodd, & Koch, 1992). These latter comparisons are made invariably on the basis of goodness-of-fit statistics that reveal whether or not the model describes the data. However, members of both classes of models can describe a given set of data in the sense that they fit statistically; therefore, the choice between models may have to be made on other considerations. First, criteria that are deemed important for response models may be articulated; second, the degree to which the models characterize the response process itself can be identified. Then the choice of model will be based on data that have a satisfactory statistical fit and are consistent with the criteria or the response process.

The comparison between the Thurstone and Rasch models made here rests on the different criteria that each satisfies and the implied processes that each characterizes. Both classes of models satisfy important criteria that are incompatible. Specifically, in the Thurstone models if a threshold is removed between a pair of categories to form a new category, then the probability of a response in the new category is the sum of the probabilities of the responses in the original categories; however, the model does not have, explicitly, the property that the estimate of the location of the person is independent of the number and location of the thresholds. For the Rasch models, if a threshold is removed before the data are collected, then the estimate of the person's location is independent of the number and location of the thresholds, but the probability of a response in the new category is not the sum of the probabilities of the responses in the original categories. Furthermore, if data satisfy the model and adjacent categories are joined after they are collected, then they will no longer satisfy the Rasch model with the smaller number of categories.

In both cases, these are properties of the models, and if the data conform to one or other of the models, then it is also a property of the data. Whether or not the data fit a model is an empirical question. This distinction between the property of the model and property of data is assumed here; therefore, reference will be restricted to properties of the models themselves.

The joining assumption. A focus for studying the properties described above, and their implications, is the *joining assumption*, defined by the property that the probability of a response in Category c or Category d is equal to the probability of responding with Category c' if Category c' replaces Categories c and d (Jansen & Roskam, 1986, p. 73).

For example, suppose a set of performances is to be graded into four graded categories—a, b, c, and d—and that each of these is defined operationally. Then any particular performance would have a probability of being classified in each of the four categories. Now suppose that instead of four categories, Categories

a and b retain the same operational definition, but Categories c and d are combined to form a new category, c', with a commensurate operational definition that subsumes the original definitions of c and d. Then if the joining assumption holds, the probability of the same performance being classified in the new Category c' is equal to the sum of the probabilities of the classifications in the original Categories c and d; furthermore, the probabilities of classifications in Categories a and b would not change.

Jansen & Roskam (1986) correctly distinguish between joining categories before and after the data are collected, and the example above illustrates joining before the data are collected. Whether or not data conform to the joining assumption is an empirical matter, but if the joining assumption is expected or required to hold, then the Thurstone model, rather than the Rasch model, is the relevant model. In the case that the categories are joined after the data are collected, the effect is analogous; that is, if the data conform to the Thurstone model before categories are joined, then they will conform with the same degree of fit to the model after they are joined; if the data conform to the Rasch model before categories are joined, then mathematically it is impossible for them to conform with the same degree of fit after the categories are joined (Rasch, 1966).

The joining assumption is intuitively appealing and leads naturally to the Thurstone model. However, its failure in the Rasch model for graded responses calls it into question and requires an explication as to what processes might lead it to hold. This paper provides such an explication, and demonstrates that in the usual cases in which graded responses are used, it is not expected that the joining assumption will hold, and therefore that the Rasch model, rather than the Thurstone model, is the relevant model.

Explicit Forms of the Rasch and Thurstone Models for Graded Responses

Both the Thurstone and Rasch models have been parameterized with varying numbers of parameters, but in order not to confuse the underlying processes represented by the models with the complications of the numbers of parameters, both models will be parameterized with exactly the same, and in the first instance, minimum number of parameters. These include the location of the person and thresholds that partition the continuum to form the categories. In addition, the entire discussion centers on the location of a person with one instrument—an item with graded responses.

The Thurstone Cumulative Probability Model

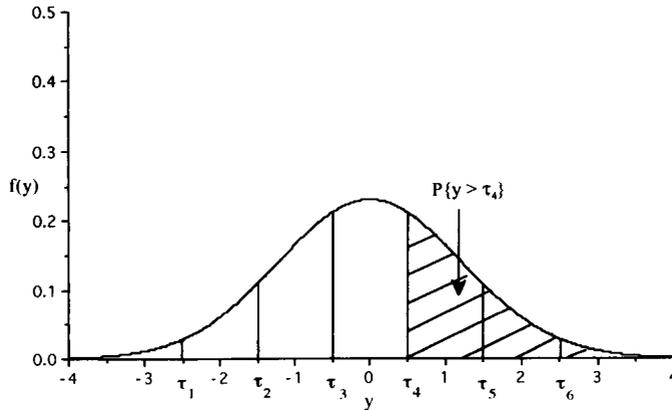
The derivation of the Thurstone model is based on the plausible assumption of a single continuous response process across a continuum, which Thurstone originally assumed to be normally distributed. If $\tau_{1i}, \tau_{2i}, \tau_{3i}, \dots, \tau_{xi}, \dots, \tau_{mi}$ are m ordered thresholds of some item i dividing the continuum into $m + 1$ categories, then the person is classified into a category depending on the realization of this process. Figure 1 shows this formulation. Because it is much more tractable, and because with a scaling constant it is virtually indistinguishable from the normal distribution (Bock, 1975), it is now often assumed that the response process is given by the double exponential (density of the cumulative logistic) distribution, $f(y) = [\exp(y)]/[1+\exp(y)]^2$, rather than the normal distribution. Thus, if Y_{pi} is a random continuous process on the continuum about the location β_p of person p , and if successive categories of item i are denoted by successive integers x_{pi} , then an outcome $\tau_{xi} \leq y_{pi} \leq \tau_{(x+1)i}$ leads to the outcome x_{pi} , with $x_{pi} = 0$ if $y_{pi} \leq \tau_{1i}$, and $x_{pi} = m$ if $y_{pi} \geq \tau_{mi}$. Formally,

$$P\{y_{pi} > \tau_{xi}\} = \int_{\tau_{xi}}^{\infty} \frac{\exp(y_{pi} - \beta_p)}{[1 + \exp(y_{pi} - \beta_p)]^2} dy_{pi} = \frac{\exp(\beta_p - \tau_{xi})}{1 + \exp(\beta_p - \tau_{xi})} \tag{1}$$

Let

$$P^*\{x_{pi}\} = P\{x_{pi}\} + P\{(x+1)_{pi}\} + P\{(x+2)_{pi}\} + \dots + P\{m_{pi}\} \tag{2}$$

Figure 1
 A Continuous Response Process on a Partitioned Continuum According to the Thurstone Model



be the cumulative probability from category x_{pi} to the last category. Then

$$P\{y_{pi} > \tau_{ki}\} = P^*\{x_{pi}\} = \frac{\exp(\beta_p - \tau_{ki})}{1 + \exp(\beta_p - \tau_{ki})}, \tag{3}$$

where $x_{pi} \in \{0, 1, 2, \dots, m\}$ is the count of the category in which the person is located.

Further, the probability of each value of the outcome x_{pi} is given by the difference of successive cumulative probabilities

$$P\{x_{pi}\} = P^*\{x_{pi}\} - P^*\{(x+1)_{pi}\}, \quad x = 1, \dots, m-1; \tag{4}$$

that is,

$$P\{x_{pi}\} = \frac{\exp(\beta_p - \tau_{xi})}{1 + \exp(\beta_p - \tau_{xi})} - \frac{\exp(\beta_p - \tau_{(x+1)i})}{1 + \exp(\beta_p - \tau_{(x+1)i})}, \quad x = 1, \dots, m-1, \tag{5}$$

and

$$P\{0_{pi}\} = 1 - \frac{\exp(\beta_p - \tau_{li})}{1 + \exp(\beta_p - \tau_{li})}, \quad P\{m_{pi}\} = \frac{\exp(\beta_p - \tau_{mi})}{1 + \exp(\beta_p - \tau_{mi})}. \tag{6}$$

This is the specific form of the Thurstone class of models that will be juxtaposed with the Rasch class of models. Because the core of the model is structured according to cumulative probabilities, and the probability in any category is the difference between successive cumulative probabilities, it is evident that the probability in any category depends only on the locations of the thresholds bounding that category. Furthermore, it is evident that, for the same reason, if a threshold is removed, the joining assumption holds immediately. It is reiterated that this is a property of the model that holds before or after any data are collected. Whether or not the property holds in data is an empirical matter, and if it does not, then the model is not relevant for those data.

The Rasch Cumulative Threshold Model

In contrast to a plausible assumption described essentially pictorially and then expressed as a model by

Thurstone, the Rasch model is based on explicit requirements of the invariance of comparisons, which then is rendered mathematically and specialized to the case of graded responses. Rasch (1961) expressed his requirements as follows:

The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; ...

Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for comparison; ... (p. 322)

Rasch termed such comparisons to be *specifically objective—objective* because they were invariant, and *specifically objective* because they were invariant within a specified frame of reference.

The necessary and sufficient mathematical expression that satisfies these requirements for dichotomous responses in a discrete stochastic context can be expressed in a number of forms all equivalent to

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \exp[(\beta_p - \delta_i) x_{pi}], \quad -\infty < \beta_p < \infty, \quad -\infty < \delta_i < \infty, \quad (7)$$

where

β_p is again the location parameter of person p ,

δ_i is the location parameter of item i , often termed its threshold, X_{pi} ,

$x_{pi} \in \{0, 1\}$ is the random variable characterizing the responses, and

$$\gamma_{pi} = \sum_{x=0}^1 \exp[(\beta_p - \delta_i) x_{pi}] \quad (8)$$

is a normalizing factor ensuring $0 < P\{x_{pi}\} < 1$ (Fischer, 1981; Rasch, 1961, 1968; Wright, 1985). This normalizing factor is always the sum of the numerators in the Rasch models.

The generalization of Equation 7 to the case of $m + 1$ ordered response categories also can be expressed in different ways, all of which are equivalent to, or restricted versions of,

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \exp[\kappa_{xi} + x_{pi}(\beta_p - \delta_i)], \quad (9)$$

where

$$\kappa_{xi} = -\sum_{k=1}^x \tau_{ki}, \quad (10)$$

τ_{ki} , $k = 1, \dots, m$, are m ordered thresholds [$\tau_{(x+1)i} > \tau_{xi}$] of item i on the latent continuum, and in which, without loss of generality,

$$\sum_{k=1}^x \tau_{ki} = 0, \quad (11)$$

and $x_{pi} \in \{0, 1, 2, \dots, m\}$ and β_p are defined as in the Thurstone model (Andrich, 1978). Again without loss of generality, the location parameter δ_i can be absorbed into the threshold parameters according to

$$\begin{aligned} \kappa_{xi} + x(\beta_p - \delta_i) &= -\tau_{1i} - \tau_{2i} \dots - \tau_{xi} + x(\beta_p - \delta_i) = -(\tau_{1i} + \delta_i) - (\tau_{2i} + \delta_i) + \dots - (\tau_{xi} + \delta_i) + x\beta_p \\ &= -\tau_{1i}^* - \tau_{2i}^* \dots - \tau_{xi}^* + x\beta_p, \end{aligned} \quad (12)$$

thereby eliminating the item parameter δ_i . In this case, of course, because δ_i has been added to each threshold of the set that sums to 0, the sum of the new thresholds is

$$\sum_{x=1}^m \tau_{xi}^* \neq 0. \quad (13)$$

The notation (*) may be dropped, because it is understood that the new thresholds have the location parameter of the item absorbed in them, giving the simpler expression

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \exp(-\tau_{1i} - \tau_{2i} \dots - \tau_{xi} + x_{pi}\beta_p) = \frac{1}{\gamma_{pi}} \exp(\kappa_{xi} + x_{pi}\beta_p). \tag{14}$$

For completeness of the expression in Equation 14, $\kappa_{0i} = \tau_{0i} \equiv 0$. Because it involves only the location of person p and the thresholds τ_{ki} , $k = 1, \dots, m$, of item i that partition the continuum into $m + 1$ ordered categories, this form of the model (Masters, 1982) makes the interpretation of the model simpler for the purpose of this paper. This form of the Rasch model (i.e., Equation 14) also gives a number of parameters identical to the form of the Thurstone model expressed in Equation 5, and therefore permits a ready comparison of the criteria and processes characterized by the models. The model of Equation 14 will be juxtaposed to the Thurstone model of Equations 5 and 6. In contrast to the Thurstone model—the core of which is the cumulative probability structure across thresholds—by the successive sum of thresholds from the first to the x th threshold, the core of the Rasch model is a cumulative threshold structure evident in the exponent of Equation 14.

In this, and all Rasch models, sufficient statistics for the parameters play a central role—it is through such statistics that one set of parameters can be eliminated and the other set estimated (Andersen, 1977; Fischer, 1981; Rasch, 1977; Wright, 1985). For example, if person p responds to items i and j , and the total score is given by $r_p = x_{pi} + x_{pj}$, then the probability of the response pair (x_{pi}, x_{pj}) , given the total score of r_p , is given by

$$P\{(x_{pi}, x_{pj})|r_p\} = \frac{\exp(\kappa_{xi} + \kappa_{xj})}{\gamma_{rij}}, \tag{15}$$

where

$$\gamma_{rij} = \sum_{(x_{pi}, x_{pj})|r} \exp(\kappa_{xi} + \kappa_{xj}), \tag{16}$$

and where

$$\sum_{(x_{pi}, x_{pj})|r} \tag{17}$$

denotes a summation over all possible pairs (x_{pi}, x_{pj}) that have a total score of r_p . This conditional probability, which can be used to construct a conditional likelihood equation for estimating the item parameters, is independent of the person location β_p . The elimination of the person parameters in the estimation of the item parameters is now so well documented (Andersen, 1977; Fischer, 1981; Rasch, 1977; Wright, 1985) that it is not explained further here; however, it is illustrated in a special case below. Figure 2 shows the probability response functions for each of four ordered response categories according to Equation 14.

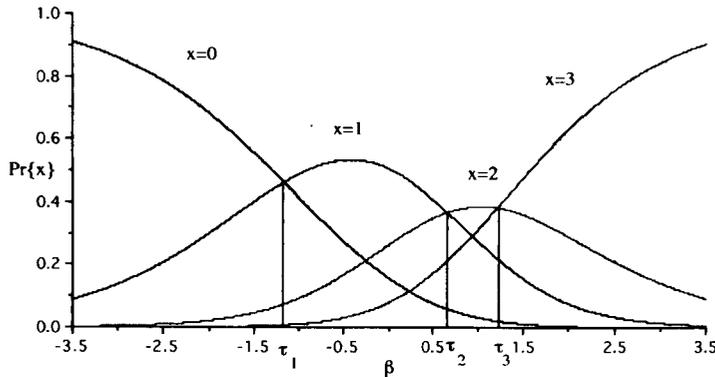
Although the model has the property that one set of parameters can be eliminated while the other set is estimated, as already stated, the model does not satisfy the joining assumption (Rasch, 1966). That is, in the notation of Equation 14, if two adjacent categories denoted x and $x+1$ are joined to give a new category x' , the probability of the response in the new category cannot be written in the form of the Rasch model. Specifically,

$$P\{x_{pi}\} + P\{(x_{pi} + 1)\} = \frac{1}{\gamma_{pi}} \exp(\kappa_{xi} + x\beta_p) + \frac{1}{\gamma_{pi}} \exp[\kappa_{(x+1)i} + (x+1)\beta_p] \neq \frac{1}{\gamma_{pi}} \exp(\kappa'_{x'i} + x'_{pi}\beta_p) \tag{18}$$

for any values of $\kappa'_{x'i}$ and β_p except in certain degenerate cases (Andersen, 1977).

These degenerate cases occur when the discrimination at a threshold is 0, in which case the response on either side of the threshold is random (Andrich, 1978). Jansen & Roskam (1986) provided a formal proof of

Figure 2
Probability Response Functions for Four Graded Responses According to the Rasch Model



this result (a brief illustration of the result is provided in the Appendix here). Furthermore, it is evident from Equation 9, in which the denominator contains all of the thresholds, that the probability of an observation in any category depends on the locations of all thresholds, not just the thresholds bounding a category. Thus, the addition or elimination of a discriminating threshold affects the distribution of responses in all categories, not just the categories adjoining the threshold. The breakdown of the joining assumption is discussed again below.

A Special Case of the Rasch Model

Historically, the Thurstone model appeared before the Rasch model. Because it is a mathematical formulation of a plausible pictorial representation, the Thurstone model is simpler intuitively than the Rasch model, which follows from abstract requirements. However, the consequences of the Rasch model challenge the joining assumption satisfied by the Thurstone model. When considered more closely, this breakdown of the joining assumption in the Rasch model clarifies the role of the instrument in generating observed responses, and therefore it is now considered in detail. In studying it in greater detail, however, it becomes instructive to consider a further special case of the model.

The Rasch Model in the Multiplicative Form

It is common in the social sciences to express parameters in models such as the Rasch model in the additive logarithmic form. In contrast, it is common in the physical sciences to express them multiplicatively (Ramsay, 1975). The Rasch model can be expressed readily in both forms, and Rasch (1960) did so. Therefore, in part because the expression of the model and the derivations to be carried out are simpler in the multiplicative form, and in part because the interpretation can be made to parallel the prototype of measurement more closely (thereby clarifying the role of an instrument), the multiplicative form of the model is used here.

To obtain the multiplicative form, in Equation 14 let $\xi_p = \exp\beta_p$ and $\omega_{xi} = \exp\tau_{xi}$. In this case, $\xi_p > 0$, $\omega_{xi} > 0$ (where $x = 0, \dots, m$), and $\omega_{0i} \equiv 1$ giving the parameters a natural origin of 0. Furthermore, because the logarithmic transformation is monotonic, $\omega_{(x+1)i} > \omega_{xi}$, and the thresholds of the continuum can be mapped and interpreted readily on a linear continuum. Inserting these transformations in Equation 14 gives

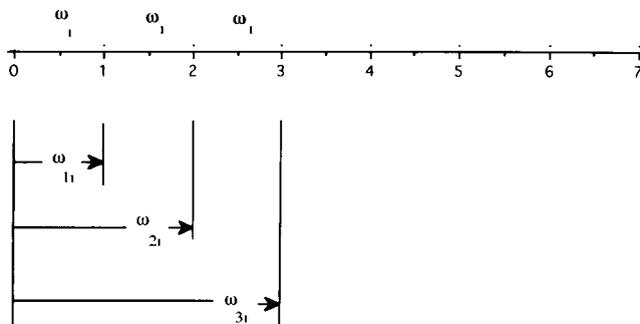
$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \left(\frac{\xi_p^{x_{pi}}}{\omega_{0i}\omega_{1i}\omega_{2i}\dots\omega_{xi}} \right), \quad \xi_p > 0; \quad \omega_{xi} > 0. \tag{19}$$

Note that in the structure of the model, the integer random variable X_{pi} , $x_{pi} \in \{0, 1, 2, \dots, m\}$, is the count or number of thresholds exceeded from the origin (Andrich, 1978), and that this integer variable does not depend on the thresholds being equidistant. Any differences in the distances between the thresholds is accounted for by their estimates.

Equal units. For the purpose of developing the special case of the Rasch model, however, suppose that the distances between thresholds are in fact equal. This specification parallels that found in the prototype of measurement in which the common distance between thresholds (including the distance between the origin and the first threshold) is termed a unit. Figure 3 shows the continuum partitioned in this way with ω_i being defined as the unit of item i , that is,

$$\omega_{0i} \equiv 1, \quad \omega_{1i} = \omega_i, \quad \omega_{2i} = 2\omega_i, \quad \omega_{3i} = 3\omega_i, \dots, \omega_{xi} = x\omega_i, \dots, \omega_{mi} = m\omega_i. \quad (20)$$

Figure 3
 A Continuum Partitioned Into Equal Units From a Natural Origin



This specification simplifies Equation 19 to

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \left(\frac{\xi_p^{x_{pi}}}{1 \cdot \omega_i 2\omega_i 3\omega_i \dots x_{pi} \omega_i} \right) = \frac{1}{\gamma_{pi}} \left(\frac{\xi_p^{x_{pi}}}{1 \cdot 1 \cdot 2 \cdot 3 \cdot \dots \cdot x_{pi} \omega_i^{x_{pi}}} \right), \quad (21)$$

which can be simplified further to

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \frac{(\xi_p / \omega_i)^{x_{pi}}}{1 \cdot 1 \cdot 2 \cdot 3 \cdot \dots \cdot x_{pi}}. \quad (22)$$

Unbounded number of units. Next, suppose that, as in the prototype of measurement, the number of categories is not a finite number fixed in advance, but that in principle it is determined by the amount of the variable being measured. If the number of categories in the system is not finite, then $m \rightarrow \infty$. Therefore, m is replaced by ∞ and $1 \times 2 \times 3 \times \dots \times x$ is replaced by $x!$ in Equation 22, which gives

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \frac{(\xi_p / \omega_i)^{x_{pi}}}{x_{pi}!}, \quad (23)$$

where the normalizing factor γ_{pi} can be simplified to

$$\gamma_{pi} = \sum_{x_{pi}=0}^{\infty} (\xi_p / \omega_i)^{x_{pi}} / x_{pi}! = \exp(\xi_p / \omega_i). \quad (24)$$

Inserting the explicit form of the normalizing factor of Equation 24 into Equation 23 gives, in the conventional mode of expression, the Poisson distribution

$$P\{x_{pi}\} = \exp(-\xi_p/\omega_i) \frac{(\xi_p/\omega_i)^{x_{pi}}}{x_{pi}!} \tag{25}$$

The explicit Poisson form, which arises when the Rasch model is expressed multiplicatively and the categories are of equal size and unbounded, is particularly convenient for analyzing the effect of changing the size and number of categories and for appreciating the role of the instrument in governing the manifest responses. It is convenient because the expressions for the expected value and variance of the Poisson distribution are simple—specifically, they are equal and identical to the parameter of the model; that is, in operator notation,

$$E[X_{pi}] = V[X_{pi}] = \xi_p/\omega_i \tag{26}$$

The Distribution of Responses as a Function of the Size of the Categories

Recall that the unit ω_i is the distance between two successive thresholds and therefore it is the size of each response category on the latent continuum. However, to simplify the discussion, this common size of the categories will continue to be referred to as the unit (as in the prototype of measurement). To appreciate the effect of changing the size of the categories by changing the location of the thresholds, suppose that a new threshold is inserted between each existing threshold; that is, a new unit, one-half the size of the previous unit, is formed. This is equivalent to halving the length of each category on the continuum. According to the joining assumption, the probability of the responses in the two halves of any category is equal to the probability of the response in the original category. The following analysis explains why the joining assumption breaks down in the Poisson model of Equation 25, and therefore also leads to an understanding of why it breaks down in the Rasch model of Equation 14.

In order to formalize the relationship between the original and new units, let $\omega_1 = 1$ be the size of the standard unit for item i ; let $\omega_2 = \omega_1/2$ be the size of the new unit that is half of the standard unit; and let X_{p2} , $x_{p2} \in \{0, 1, 2, \dots\}$, be the new integer random variable in the half units. Then $X_{p1} = X_{p2}/2$. For example, if ω_1 is a centimeter, then X_{p1} is a count of the number of centimeters exceeded from the origin, and X_{p2} is the count of the number of half-centimeters exceeded from the origin. Let $X_{p2}^{(1)}$, $x_{p2}^{(1)} \in \{0/2, 1/2, 2/2 \dots\}$ be the rational random variable of measurements in the unit ω_2 , but expressed in units ω_1 . For example, if measurements are taken in half-centimeters but expressed in centimeters, the possible values are 0/2, 1/2, 2/2, 3/2,

Estimate of ξ_p . In the first step in the analysis of the consequence of halving the size of the unit, consider the effect on the estimate of ξ_p .

According to Equation 25, in the standard unit $\omega_1 = 1$,

$$P\{x_{p1}\} = \left[\exp(-\xi_p/\omega_1) \right] / x_{p1}! \tag{27}$$

and

$$E\{X_{p1}\} = \xi_p/\omega_1 = \xi_p/1 = \xi_p \tag{28}$$

Thus, an unbiased estimate of ξ_p is simply the count of the number of standard units exceeded from the origin, which is analogous to the prototype of measurement.

Now consider the case when the unit is halved so that $\omega_2 = \omega_1/2$. From Equation 25 again

$$P\{x_{p2}\} = \left[\exp(-\xi_p/\omega_2) \right] / x_{p2}! \tag{29}$$

and

$$E[X_{p2}] = \xi_p / \omega_2 = \xi_p / (\omega_1 / 2) = 2\xi_p / \omega_1 = 2\xi_p / 1 = 2\xi_p. \quad (30)$$

Thus, in the units that are one-half the standard unit, the location of the entity is estimated to be two times that when estimated in the standard unit; this is consistent with intuition. For example, if an entity was estimated to be 10 cm long, it would be expected to be $2(10) = 20$ half-cm long.

However, to compare this estimate in the half-unit with that in the standard unit, it is necessary to express a measurement in the half-unit in terms of the standard unit. Expressing this measurement, X_{p2} , taken in the new unit ω_2 , in terms of the standard unit ω_1 , gives $X_{p2}^{(1)} = X_{p2} / 2$.

Therefore,

$$E[X_{p2}^{(1)}] = E[X_{p2} / 2] = \frac{1}{2} E[X_{p2}] = \frac{1}{2} 2\xi_p = \xi_p. \quad (31)$$

Thus, when the measurement is taken in a unit one-half the standard unit, and is expressed in the standard unit, the expected value remains the same as when the measurement is taken in the standard unit. Again, this is consistent with intuition. It is important to stress that with a different partitioning of the continuum, the estimate of ξ_p remains the same.

It is also important to appreciate the two distinctive steps in the derivation of $E[X_{p2}^{(1)}]$. In the first step, the unit is changed in the measurement process and therefore the parameter of the Poisson distribution changes from ξ_p / ω_1 to ξ_p / ω_2 . In the second step, the unit in which the measurement is expressed is changed from X_{p2} to $X_{p2}^{(1)} = X_{p2} / 2$, but the parameter of the distribution is not changed.

Precision of the estimate of ξ_p . The convenience of having the Poisson distribution as the special case of the Rasch model for graded responses should already be evident—the observed value of the random variable immediately gives an unbiased estimate of the location. The identity of the estimate of ξ_p in the standard and new units (when expressed in the standard unit) is consistent with intuition. Using the same procedure, it is possible to derive the precision of this estimate equally conveniently.

First, according to Equation 25 and in the standard unit $\omega_1 = 1$,

$$V[X_{pi}] = \xi_p / \omega_1 = \xi_p / 1 = \xi_p. \quad (32)$$

Now again consider the case where the unit is halved so that $\omega_2 = \omega_1 / 2$. In this case,

$$V[X_{p2}] = \xi_p / \omega_2 = \xi_p / (\omega_1 / 2) = 2\xi_p / \omega_1 = 2\xi_p / 1 = 2\xi_p. \quad (33)$$

Thus, in the new unit, the variance of the random variable is twice as great as in the original unit. Because X_{p2} is an unbiased estimate of ξ_p , this, in fact, is the variance of the estimate. However, to make a meaningful comparison between the precision of measurement in the standard unit and in the new unit, it is necessary again to express the measurement using the new unit in terms of the standard units. This is given by

$$V[X_{p2}^{(1)}] = V[X_{p2} / 2] = \frac{1}{4} V[X_{p2}] = \frac{1}{4} 2\xi_p / \omega_1 = \frac{1}{2} \xi_p / \omega_1 = \frac{1}{2} \xi_p / 1 = \frac{1}{2} \xi_p. \quad (34)$$

Thus when the unit is reduced by one-half, the variance of the estimate is also reduced by one-half, which means that the precision is increased. This, too, is consistent with intuition—the point of measuring in a smaller unit is to have more precise estimates of the location of the entity of measurement.

Again, the convenience of the special case of the Poisson distribution is evident in that the variance of the estimate has a single explicit expression. Furthermore, the result can be generalized to the reduction of a standard unit by any factor. For example, if the new unit is $1/n$ th the size of the standard unit, then when expressed in terms of the standard unit,

$$E[X_{pn}^{(1)}] = \xi_p \tag{35}$$

and

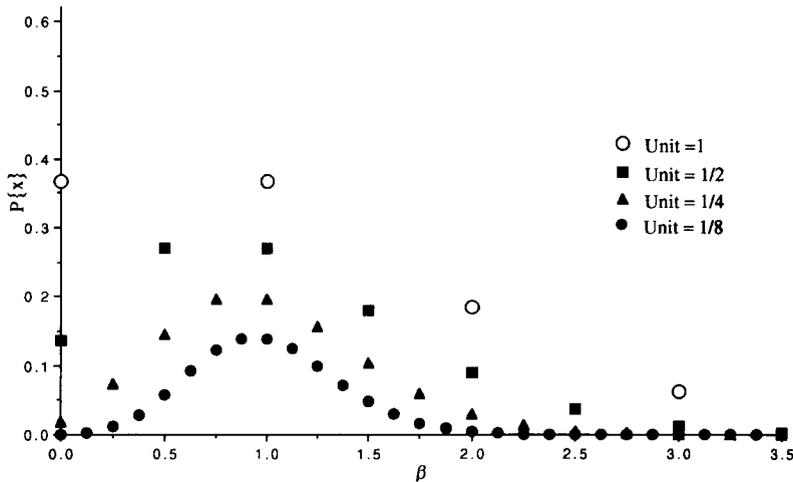
$$V[X_{pn}^{(1)}] = \frac{1}{n} \xi_p, \tag{36}$$

showing that the variance of the estimate of ξ_p reduces proportionately by $1/n$.

Figure 4 shows the distribution of the measurements when the units are divided successively into 1/2, 1/4, and 1/8 of the standard unit and the person being measured has a value of $\xi_p = 1$. From the distributions in Figure 4 it should be evident why the joining assumption cannot hold: When the continuum is partitioned into smaller units, the variance of the observed distribution changes—it becomes relatively smaller when expressed in the standard units. As a result, the sum of the probabilities of outcomes in the two adjacent smaller segments of the continuum is not the same as the probability of an outcome in the segment composed by joining the two smaller segments.

Figure 4

Distribution of Measurements for a Person Whose Latent Value is 1 Using Instruments With Units of 1/2, 1/4, and 1/8



Although the demonstration of how the joining assumption fails to hold is clear in the special case of the Poisson distribution when the Rasch model is expressed multiplicatively and the distances between an unbounded number of thresholds are equal, the same feature holds when the thresholds are finite in number and not equidistant, and it holds for the same reason (i.e., as the number of thresholds is increased, the precision of the estimate of the location of the entity is also increased) (Andrich, in press).

Abstraction of the Role of the Instrument in Manifesting Response Distributions

To consolidate the way the precision of the estimate affects—and destroys—the joining assumption, and to lead into an analysis of the Thurstone model for graded responses, another common example is now examined briefly. Suppose, as in classical test theory (CTT; e.g., Gulliksen, 1950; Lord & Novick, 1968), that there is a distribution of locations β of persons of a population, say $\beta \approx N(\mu, \sigma_\beta^2)$. Suppose, further, that each person p is tested with some instrument j to give an observed score T_{pj} for each person, where this score contains an error e_{pj} , giving $T_{pj} = \beta_p + e_{pj}$. (To be consistent with the latent trait formulation, β is used

for the true score usually denoted by T in classical test theory.) Then, if the errors are distributed according to $e_j \approx N(0, \sigma_j^2)$ and are uncorrelated with the latent locations' β ,

$$V[T_j] = V[\beta] + V[e_j], \tag{37}$$

or if $\sigma_{T_j}^2 = V[T_j]$,

$$\sigma_{T_j}^2 = \sigma_\beta^2 + \sigma_j^2. \tag{38}$$

Figure 5 shows the distribution of T_j . Suppose now a second instrument k is used with errors $e_k \approx N(0, \sigma_k^2)$, but where $\sigma_k^2 > \sigma_j^2$. In conventional CTT terms, the reliability of instrument k is lower than that of instrument j and its measurements have lower precision. Then

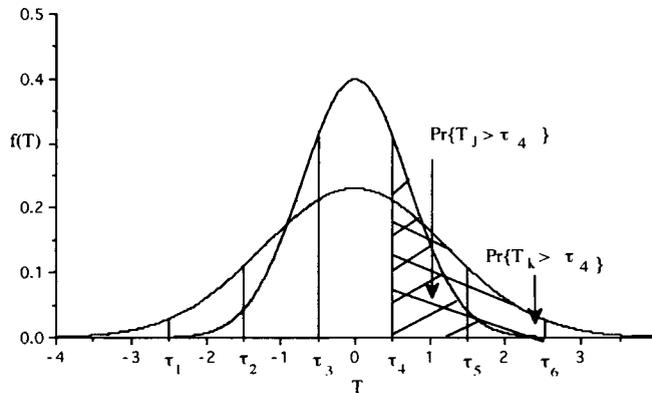
$$V[T_k] = V[\beta] + V[e_k]. \tag{39}$$

That is,

$$\sigma_{T_k}^2 = \sigma_\beta^2 + \sigma_k^2 > \sigma_{T_j}^2. \tag{40}$$

Figure 5 also shows the distribution of T_k . In addition to the distributions of T_j and T_k , Figure 5 shows a series of thresholds that divide the distribution in accordance with the rationale for the Thurstone model shown in Figure 1.

Figure 5
Two Continuous Response Processes on a Partitioned Continuum According to Classical Test Theory



It should now be evident that in moving from instrument j to instrument k in the same population, the probability of an observed score T_p between any pair of thresholds changes. The joining assumption now does not hold when using one instrument with one precision rather than another instrument with another precision because the probability of a score being between any pair of thresholds has changed. The point of this illustration is to show that it is exactly the kind of situation characterized by the Rasch model—if a threshold is added or removed, it produces a new instrument with a different precision in collecting data; as a consequence, the variance of the observations changes and the joining assumption no longer holds.

The explication of how the joining assumption fails in the Rasch model and in CTT when using different instruments with different precision of measurement raises the question as to which conditions would permit the joining assumption to hold under a change of instruments. In the case of CTT, it would hold if the precision of the two instruments were identical, including the special condition in which there is no error of

measurement in either instrument. In the Rasch model, which is sensitive to the number and location of the thresholds, the joining assumption can only hold, as has already been indicated, if the discrimination at a threshold is 0. This is consistent with the above interpretation in terms of precision of measurement because if the threshold that is removed does not discriminate, it contributes nothing to improving the precision of measurement.

Re-examination of the Thurstone Model

To anticipate an important contrast between the Rasch model and the Thurstone model, the invariance of the estimate of the location of the person parameter in the Rasch model is highlighted. It was shown in the analysis of the special case of the Rasch model that resulted in the Poisson distribution that the estimate of ξ_p itself was invariant under different partitions of the continuum; moreover, it was shown that as the number of thresholds increased, the precision of the estimate of a location parameter also increased. It can be shown that both of these features hold for the general Rasch model of Equation 14 for ordered response categories (Andrich, in press).

Given these features of the Rasch model and the CTT analysis of the role of error, consider once again the Thurstone model for ordered response categories. First, because the Thurstone models do not have sufficient statistics, the person parameter cannot be eliminated while the threshold parameters across items are estimated, or vice versa as was done for the Rasch models. This means that, in principle, the estimate ξ_p is not independent of the location of the thresholds. Second, the joining assumption holds for the Thurstone models; furthermore, if any threshold is added or removed, the probability of the responses in all other categories, except those adjoining the threshold that was added or removed, remains the same. This means that no matter how many thresholds are added to the instrument, the variance of the response process shown in Figure 1 remains the same. Therefore, this variance cannot be attributed to the instrument as a function of the number and location of the thresholds (as in the Rasch model)—it is all attributed to the person, even though the person has only one parameter ξ_p . This has major implications for the viability of the model for the same situation for which the Rasch model might be appropriate, and vice versa, as discussed above.

To further develop the case for the Thurstone model, it is instructive to consider an example in which the joining assumption would hold. Suppose that the entity of measurement is a population of individuals, each with a location, and that this population has a mean and a variance, just as in the CTT analysis. At this point in the analysis, it is necessary to distinguish between an observed (manifest) location of an entity on a continuum and the recorded response. For example, suppose measurements are to be recorded in centimeters so that the thresholds are 1 cm apart. A measurement of the length of some object is recorded as 10 cm to the nearest centimeter if the entity is deemed to be anywhere between $10 \pm .5$ cm. Where the object is deemed to be observed (e.g., at 10.05 cm) is termed the observed location, and the recorded response then is 10 cm. Now if the observed location of a person was identical no matter which instrument was used (i.e., the number and location of the thresholds played no role in producing the manifest locations), then the joining assumption would hold; that is, to be specific, if the continuum were repartitioned by thresholds that were 2 cm apart rather than 1 cm apart, then the probability of an observed location of a randomly selected individual being in category x' (a 2 cm range), where category x' replaces adjacent categories x and $x + 1$ (each of 1 cm range), would be the sum of the probabilities of locating this individual in categories x and $x + 1$. (This may occur, for example, in a study of the distribution of heights in which the measurement of the height of each person is sufficiently accurate that any variation in the observed location of any person's height, when the person is measured more than once, can be ignored.)

However, intuitively it would be reasonable to argue that an instrument with more thresholds nevertheless operates in some sense more precisely than an instrument with fewer thresholds. Therefore, in view of

the analysis with the Rasch model, it might be expected that the joining assumption would not hold. Thus, there is a paradox. As usual in a paradox, some ingredient in the chain of reasoning is omitted. In the case of precision in the Thurstone model, the ingredient concerns the interpretation given to the manifest location of each person and to a distinction between this location and an estimate of the mean location of the population. Before proceeding to explicate this paradox, note that considering a population in relation to an interpretation of the Thurstone model is consistent with Thurstone's formulation—he always dealt with populations of individuals and not individuals as such (Thurstone, 1925, 1928).

In order to understand the kind of increased precision more thresholds provide in the Thurstone model, consider again the population of individuals that is distributed on a continuum, where each individual is located on the continuum without any error (i.e., in an identical location from replication to replication). The location of the individuals is fixed and does not depend on the instrument. Further, suppose that the thresholds are placed on the continuum on which the distribution of individuals is displayed and are independent of the instrument that gave rise to the measurements. Then if the thresholds are close together, the recorded location of each individual of the population will be more accurate than if the thresholds are further apart. For example, if the thresholds are 1 cm apart, the recorded location of each individual will be more accurate than if the thresholds are located 2 cm apart.

Here lies the resolution of the paradox highlighted above of how in the Thurstone model the joining assumption holds, yet at the same time there is an increase in precision if the number of thresholds is increased. The increase in precision pertains to the recording of the location of each individual, but this increase in the precision of recording the location of each individual has only an indirect bearing on the precision of the estimate of the location of the mean of the population of individuals based on a random sample of individuals. Indeed, it is well known from traditional textbooks (e.g., Guilford, 1965, p. 85) that when class intervals are formed after data are collected, and the variance is calculated based on the central value of each class interval, there is no bias in the estimate of the mean but there is bias in the variance estimate, and that this bias can be corrected. The correction is intended to make the variance independent of the size of the class intervals.

Thus, the case in which the joining assumption does hold involves an entity (e.g., a population of individuals) that is characterized by both a mean and a variance in which the observed location of each manifestation of the population is not affected by the measuring instrument. In extrapolating this example to the case of the response of a single person, it might be possible to characterize a single person with both a mean and a variance. For example, a person may be characterized by a location on a continuum and a latitude of acceptance (Coombs, 1964). However, in the models for graded responses in item response theory (IRT) this generally is not done. Instead, the Thurstone and Rasch models are both parameterized to include only the location parameter for the individual person, as in Equations 5 and 9, respectively, and threshold parameters for items. Moreover, if a person were characterized by both a mean and a variance, and the joining assumption did hold, then the observed location of the individual on any replication would be more precise with thresholds close together than far apart, but that would not provide a better estimate of the single (latent) location parameter of the person.

For the terminology in which the observed (manifest) location is distinguished from the recorded response, the conclusion that follows from the above analysis is that the Thurstone model, in which the joining assumption holds, is appropriate only when the distribution of the observed locations is entirely due to the person and when the location of the thresholds plays no role in manifesting actual observed locations. However, given the observed location of a person on any replication, the location of the thresholds governs how precisely that observed location is recorded. Accordingly, the usual use of the Thurstone models for graded responses in which only the location of the person is parameterized and the distribution of responses is presumed to be governed by the instrument and the location of its thresholds, cannot be

supported—to put it simply, it is internally inconsistent.

Rasch Models Versus Thurstone Models

The distinctive features of the processes characterized by the Rasch and Thurstone models when applied to graded responses in IRT are as follows. In the Rasch model (1) the person has a single location; (2) the source of the final distribution of the observed locations resides entirely in the instrument; (3) the probability of a response in any category depends on the location of all thresholds, not just the locations of the thresholds bounding the category; (4) the joining assumption does not hold; (5) the variance of this distribution is inversely related to the number of thresholds; (6) the precision of an estimate of the single location increases with an increase in the number of thresholds; and (7) the estimate of the location of the person can be separated explicitly from the location of the thresholds.

In the Thurstone model (1) the person also has a single location; but (2) the source of the distribution of the observed locations resides entirely in the person; (3) the probability of a response in a category depends on the location of only the thresholds bounding the category; (4) the joining assumption holds; (5) the variance of this distribution is unrelated to the number of thresholds; (6) the precision of the location of each realized observed location in the empirical distribution increases with the number of thresholds, but the precision of the estimate of the single location parameter of the person is not improved; and (7) the estimate of the single location of the person cannot be separated explicitly from the location of the thresholds.

It seems from this juxtaposition that the essential difference between the processes characterized by the two models is that in the Rasch model each replicated observation is considered incidental and transient, with the main goal being the precision of the estimate of the single latent location ξ_p of individual p . By contrast, in the Thurstone model, each observation is treated as an end in itself and the main goal is to locate each observation accurately.

Further Implications

A Restriction On the Use of the Thurstone Model

The above analysis of the essential properties of the two generic models for graded responses showed that they have incompatible properties; therefore, they should not be treated as alternative models for the same kind of data in which the choice of which model to use is simply a matter of preference. Nevertheless, if the two models were applied to a particular dataset, it is likely that if one model accounted for the data in a statistical sense, the other would likely do so as well. Therefore, alternative approaches need to be devised to establish which, if either, of the models characterizes not just the data, but the process underlying the data.

One approach is to conduct experiments directed by the model. For example, Andrich (1993, in press) reported two studies, one based on the work of Harris (1992), in which the same people were required to produce two sets of graded responses when encountering the same stimulus. There were three categories in a set of graded responses in one case and four in the other case. Furthermore, in each study, the fourth category was a subset of the third category that was intended to maximize the opportunity for the joining assumption to hold. In both cases the joining assumption did not hold; that is, when the extra category was introduced, the probabilities of all other categories changed, not just the probability of the third category. In these cases, therefore, the Rasch model rather than the Thurstone model was more appropriate. It seems from the discussion in the previous section and these examples that this would hold for all Likert-type scales and, therefore, that the Thurstone models might not be appropriate for such response formats. Whether this is true in any specific case is an empirical question.

A Restriction On the Use of the Rasch Model

Having concluded that there is potentially a major restriction on the use of the Thurstone model, it may be instructive to emphasize that the above analysis also highlights a major restriction, and a special interpretation, of the Rasch model. This restriction is that the model does not characterize a sequential, but a simultaneous, process across the thresholds. Because the model pertains to a continuum in which there is a preferred direction, and the ordered categories are defined on the continuum, it is tempting to think that the model characterizes a sequential process of proceeding from one category to the next in some ordered manner; for example, solving a problem in a sequence of unique well-defined steps. This is clearly not the case, as demonstrated by the feature that the probability of the response in any category depends on the location of the thresholds between all categories, not just the preceding category. In particular, the probability of the response in the first category depends on the location of the threshold between the last and second to last categories. For example, in responding to a Likert-style format in a questionnaire, a person can see all categories simultaneously and selects from among all categories, and the Rasch model seems relevant.

This means that the Rasch model cannot represent a process of going from one category to the next, where a particular response depends only on the previous observed response. Because the probability of a response in any category depends on the location of all of the thresholds, the person must have some knowledge of all thresholds—relative to any position, it requires knowing what has preceded as well as what lies ahead (Tutz, 1990). As described briefly below, this indeed is likely to be the case in many common circumstances in which the Rasch model might be applied.

For example, in producing a performance (e.g., the solution to a problem or writing an essay), the person has a conceptualization of the whole that generates the component parts as well as the sequence in which the components make up this whole; that is, in executing each component of the performance, the person has in mind all the previous components as well as the remaining components, including the final component in the sequence. The person does not simply produce a component in a sequence that depends only on the previous component. Likewise, in grading a performance, there is an expectation of which components are to be executed to make up the whole as well as the sequence (more or less correct or consistent) in which they are to be executed. This does not mean that the sequence of components is unique, but neither is it arbitrary. In some sense, this is analogous to the expression of an idea in a sentence, in which the same idea can be expressed in many different sentences, but the sequence of words and phrases in which the idea is expressed is also constrained by the rules of grammar and is not arbitrary.

Once again, the special Poisson case of the Rasch model (Equation 25) is useful in appreciating the meaning of the threshold parameters: each threshold ω_{xi} shown in Figure 3 is defined immediately from the same origin, and characterizes its distance from (or difficulty relative to) the same origin. It does not represent executing component x given that component $x - 1$ has been executed, where category x is equated here to component x ; it is simply a threshold marking off the distance from the origin that separates the components (i.e., categories in the sequence).

For example, in the dichotomous case of the Rasch model applied to an achievement item, a correct response is not thought of as an indication of going from 0 (having reached being incorrect) to 1 (being correct given that the person has reached being incorrect), but simply the difficulty of being correct relative to being incorrect. Likewise, the parameter τ_{xi} indicates the relative difficulty of being in category x rather than category $x - 1$ (given that the person is either in category $x - 1$ or category x). Thus, the Rasch model of Equation 14 can be viewed as a simultaneous, rather than a sequential, processing model. Appreciating the way the model characterizes these kinds of responses can lead to more powerful uses of the model than merely to describe the data in terms of parameters. Sequential processing models that have been related to

Rasch-type models have been studied (Kempf, 1976; Verhelst & Glas, 1991), but the Rasch model for graded responses studied in this paper is not one of them. Andrich (1995a, 1995b) and Roskam (1995) considered related properties of the Rasch model.

Joint Use of the Thurstone and Rasch Models

Having elaborated the different applications and restrictions on the use of both the Thurstone and Rasch models, it may be useful to consider how they might be used together. For example, they could be used together when the “entity” is a group of individuals and the location of each individual is obtained using a graded response format in which the thresholds play a role in the precision of the estimate of this location. Then the group of individuals would have both a mean and a variance and the joining assumption would be expected to hold at the level of the latent parameters β . Taking the logistic of the density function of the distribution of persons results in

$$P\{x_{pi} \cap \beta_p\} = \frac{1}{\gamma_{pi}} \exp(-\tau_{1i} - \tau_{2i} \dots - \tau_{xi} + x\beta_p) \left[\frac{\exp\alpha(\beta_p - \tau_{(x-1)i})}{1 + \exp\alpha(\beta_p - \tau_{(x-1)i})} - \frac{\exp\alpha(\beta_p - \tau_{xi})}{1 + \exp\alpha(\beta_p - \tau_{xi})} \right], \tag{41}$$

from which the unconditional probability of a response x_{pi} , $P\{x_{pi}\}$, for a randomly selected member of the population would be given by

$$P\{x_{pi}\} = \int \frac{1}{\beta \gamma_{pi}} \exp(-\tau_{1i} - \tau_{2i} \dots - \tau_{xi} + x\beta) \left[\frac{\exp\alpha(\beta_p - \tau_{(x-1)i})}{1 + \exp\alpha(\beta_p - \tau_{(x-1)i})} - \frac{\exp\alpha(\beta_p - \tau_{xi})}{1 + \exp\alpha(\beta_p - \tau_{xi})} \right] d\beta, \tag{42}$$

where α characterizes the dispersion of the distribution of persons. Further development of this model, however, is beyond the scope of this paper.

Conclusions

The existence of two well-developed models for graded responses has facilitated not only a detailed comparison between the models but a better understanding of each. If only one model were well-developed, then its distinctive properties could not be compared against the contrasting features of the other model. Accordingly, it is acknowledged that the analysis of each of the models in this paper has depended greatly on previous work with each model. The conclusions of this paper are surprising, but when the distinctions between the Rasch and Thurstone models are appreciated fully, they will permit a better understanding of the choice of models for graded responses and a better interpretation of the data when they are analyzed according to the chosen model.

Appendix:

**The Feasibility of Joining Adjacent Categories in the Rasch Model
When the Discrimination at the Threshold Between the Categories is 0**

In order to demonstrate that categories can be combined when the discrimination at a threshold is 0, it is necessary to introduce a discrimination parameter at each threshold. Let $\alpha_1, \alpha_2, \dots, \alpha_x, \dots, \alpha_m$ be the m successive discriminations at the m respective thresholds τ_{xi} , $x = 1, \dots, m$ (Andrich, 1978). This threshold discrimination is similar to the item discrimination at the difficulty (threshold) of a dichotomous item in the two-parameter logistic model (Birnbaum, 1968). If different threshold discriminations are incorporated in the model of Equation 14, the model takes the form (Andrich, 1978)

$$P\{x_{pi}\} = \frac{1}{\gamma_{pi}} \exp[-\alpha_{1i}\tau_{1i} - \alpha_{2i}\tau_{2i} - \alpha_{3i}\tau_{3i} \dots - \alpha_{xi}\tau_{xi} + (\alpha_{1i} + \alpha_{2i} + \alpha_{3i} \dots + \alpha_{xi})\beta_p]. \quad (43)$$

This model specializes to Equation 14 if all of the discriminations equal 1. Consider the special case in which the discrimination at Threshold 3, between Categories 2 and 3, is 0 (i.e., $\alpha_{3i} = 0$). Then

$$\begin{aligned} P\{2_{pi}\} + P\{3_{pi}\} &= \frac{1}{\gamma_{pi}} \exp[-\alpha_{1i}\tau_{1i} - \alpha_{2i}\tau_{2i} + (\alpha_{1i} + \alpha_{2i})\beta_p] \\ &\quad + \frac{1}{\gamma_{pi}} \exp[-\alpha_{1i}\tau_{1i} - \alpha_{2i}\tau_{2i} - 0\tau_{3i} + (\alpha_{1i} + \alpha_{2i} + 0)\beta_p] \\ &= \frac{2}{\gamma_{pi}} \exp[-\alpha_{1i}\tau_{1i} - \alpha_{2i}\tau_{2i} + (\alpha_{1i} + \alpha_{2i})\beta_p]. \end{aligned} \quad (44)$$

This is consistent with intuition in that if a threshold has 0 discrimination, then the response in the two adjoining categories will be random and the probability of the response in the two categories separated by the threshold will be twice the probability of the response in either category.

The formal proof of the converse is given by Jansen & Roskam (1986). For example, if the threshold between these two categories is not 0, then the only simplification possible for $P\{2_{pi}\} + P\{3_{pi}\}$ is

$$P\{2_{pi}\} + P\{3_{pi}\} = \frac{1}{\gamma_{pi}} \exp[-\alpha_{1i}\tau_{1i} - \alpha_{2i}\tau_{2i} + (\alpha_{1i} + \alpha_{2i})\beta_p] [1 + \exp(-\alpha_{3i}\tau_{3i} + \alpha_{3i})\beta_p], \quad (45)$$

which cannot be expressed in the form of Equation 14.

References

- Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*, 69–81.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 357–374.
- Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 33–80). San Francisco: Jossey-Bass.
- Andrich, D. (1993, September). *Where linguistics, physical measurement and social measurement converge*. Paper presented at the Fourth International Conference on Bibliometrics and Scientometrics, Berlin, Germany.
- Andrich, D. (in press). Theoretical and empirical evidence on the dichotomization of graded responses. In G. Engelhard & M. Wilson (Eds.), *Objective measurement: Theory into practice* (Vol. 3). Norwood NJ: Ablex.
- Andrich, D. (1995a). Models for measurement, precision and the non-dichotomization of graded responses. *Psychometrika*, *60*, 7–26.
- Andrich, D. (1995b). Further remarks on the non-dichotomization of graded responses. *Psychometrika*, *60*, 37–46.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397–479). Reading MA: Addison-Wesley.
- Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Dawes, R. M. (1972). *Fundamentals of attitude measurement*. New York: Wiley.
- De Ayala, R. J., Dodd, B. G., & Koch, W. R. (1992). A comparison of the partial credit and graded response models in computerized adaptive testing. *Applied Measurement in Education*, *5*, 17–34.
- Edwards, A. L., & Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive integers. *Psychometrika*, *17*, 169–180.
- Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, *46*, 59–77.
- Guilford, J. P. (1965). *Fundamental statistics in psychology and education* (4th ed.). New York: McGraw-Hill.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Harris, J. (1992). *Consequences for social measurement of collapsing adjacent categories with three or more*

- ordered categories. Unpublished master's of education thesis, Murdoch University, Western Australia.
- Jansen, P. G. W., & Roskam, E. E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika*, *51*, 69–91.
- Kempf, W. F. (1976). Dynamic models for the measurement of "traits" in social behavior. In W. F. Kempf & B. H. Repp (Eds.), *Some mathematical models for social psychology* (pp. 14–58). Bern: Huber.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140.
- 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.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, *42*, (Series B), 109–142.
- McCullagh, P. (1985). Statistical and scientific aspects of models for qualitative data. In P. Nijkamp, H. Lielner, & N. Wrigley (Eds), *Measuring the unmeasurable* (pp. 39–49). Dordrecht, The Netherlands: Nijhoff.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.
- Ramsay, J. O. (1975). Review of *Foundations of measurement* (Vol. I) by D. H. Krantz, R. D. Luce, P. Suppes, & A. Tversky. *Psychometrika*, *40*, 257–262.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. Expanded edition, University of Chicago Press, 1980.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, IV* (pp. 321–334). Berkeley: University of California Press.
- Rasch, G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 89–108). Chicago: Science Research Associates.
- Rasch, G. (1968). *A mathematical theory of objectivity and its consequence for model construction*. Paper presented at the European Meeting on Statistics, Econometrics and Management Science, Amsterdam, The Netherlands.
- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, *14*, 58–94.
- Roskam, E. E. (1995). Graded responses and joining categories: A rejoinder to Andrich's "Models for measurement, precision and non-dichotomization of graded responses." *Psychometrika*, *60*, 27–35.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*, 567–577.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, *16*, 433–451.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*, 39–55.
- Verhelst, N., & Glas, C. A. W. (1991). *A dynamic generalization of the Rasch model*. Arnheim, The Netherlands: CITO.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.
- Wright, B. D. (1985). Additivity in psychological measurement. In E. E. Roskam (Ed.), *Measurement and personality assessment. Selected papers. XXIII International Congress of Psychology*, *8* (pp. 101–111). Amsterdam, North Holland.

Acknowledgments

Pender Pedler, Jürgen Rost, and Irene Styles read earlier versions of this paper and provided constructive comments, as did two unknown reviewers and Fritz Drasgow.

Author's Address

Send requests for reprints or further information to David Andrich, School of Education, Murdoch University, Western Australia 6150. Internet: andrich@murdoch.edu.au.