A Hyperbolic Cosine Latent Trait Model For Unfolding Dichotomous Single-Stimulus Responses — Source link ↗

David Andrich, Guanzhong Luo

Institutions: Murdoch University, South China Normal University

Related papers:

- A theory of data

- The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude

- A General Item Response Theory Model for Unfolding Unidimensional Polytomous Responses

- A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies

- Attitudes Can Be Measured

A Hyperbolic Cosine Latent Trait Model For Unfolding Dichotomous Single-Stimulus Responses

David Andrich, Murdoch University, Australia

Guanzhong Luo, South China Normal University, Guangzhou

Social-psychological variables are typically measured using either cumulative or unfolding response processes. In the former, the greater the location of a person relative to the location of a stimulus on the continuum, the greater the probability of a positive response; in the latter, the closer the location of the person to the location of the statement, irrespective of direction, the greater the probability of a positive response. Formal probability models for these processes are, respectively, monotonically increasing and single-peaked as a function of the location of the person relative to the location of the statement. In general, these models have been considered to be independent of each other. However, if statements constructed on the basis of a cumulative model have three ordered response categories, the response function within the statement for the middle category is in fact single-peaked. Using this observation, a unidimensional model for responses to statements that have an unfolding structure was constructed from the cumulative Rasch model for ordered response categories. A location and unit of measurement parameter exist for each statement. A joint maximum likelihood estimation procedure was investigated. Analysis of a small simulation study and a small real dataset showed that the model is readily applicable. *Index terms: attitude measurement, item response theory, latent trait theory, Rasch models, unfolding data, unidimensional scaling.*

The two main response processes for constructing social-psychological variables are the cumulative and the unfolding processes. In the former, the greater the location of a person relative to the location of a stimulus on the continuum, the greater the probability of a positive response; in the latter, the closer the location of the person to the location of the stimulus, irrespective of direction, the greater the probability of a positive response. Thurstone (1927, 1828), who first formalized the construction of psychological variables with no physical counterpart, implied both processes (Andrich, 1988, 1989).

In addition to these two response models, there are also two main data collection designs. In the single-stimulus response design, the person responds directly to a statement by answering it either correctly or incorrectly, or by agreeing or disagreeing with it. Henceforth, this will be referred to as the "direct response" design. In the pair-comparison design, the person either compares two statements for their location on a continuum (irrespective of the person's own location) or selects one of two statements based on which one best reflects his or her own position. Again, both of these designs were invoked by Thurstone (1928). The stimuli used here will be considered to be statements such as those found in questionnaires used to measure attitude.

Coombs (1950) and his associates (Coombs & Avrunin, 1977; Coombs & Smith, 1973) studied the unfolding model, and Guttman (1950, 1954) studied the cumulative model. Both Coombs and Guttman restricted themselves to deterministic, and not formal probabilistic, response models. For the pair-comparison design, formal probabilistic models have been studied with both unfolding and cumulative models; for the direct response design, formal probabilistic models have been studied

primarily with cumulative models. This direction was set by Thurstone (1927) who formalized a cumulative probabilistic model in the pair-comparison design in his "law of comparative judgment" and a cumulative probabilistic model in the single-stimulus response design for achievement testing (Thurstone, 1926). However, Thurstone retained an ad hoc procedure for the unfolding model in the direct response design for the measurement of attitude and did not really consider the unfolding model in the pair-comparison design.

Probabilistic models for the cumulative models in pair-comparison designs have been studied extensively (Bock & Jones, 1968; Bradley & Terry, 1952; David, 1963; Luce, 1959; see Davidson & Farquhar, 1976, for a bibliography of approximately 350 articles). Applications of latent trait or item response theory to psychological and achievement testing have generated substantial work on cumulative probabilistic models for the direct response design.

Explicit probabilistic models based on unfolding in the pair-comparison design have been developed by Sixtl (1973), Zinnes & Griggs (1974), and Schönemann (1970), and have been further explicated and extended by Carroll (1972), Cooper & Nakanishi (1983), DeSarbo & Hoffman (1986), Jansen (1984), Poole (1984), Schönemann & Wang (1972), Tanzer (1983), Carroll & De Soete (1991), and Böckenholt & Böckenholt (1990). Until recently, the unfolding model for the single-stimulus response design, to which this paper is devoted, has appeared less often in the literature (Andrich, 1988; Davison, 1977; Heiser, 1989; Hoijtink, 1990; Post & Snijders, 1992; Van Schuur, 1984, 1988, 1989).

Cumulative and unfolding models have been developed separately or in parallel, but have never been shown to be directly related. Here, the development of a model for unfolding is derived explicitly from a cumulative model. Specifically, when the cumulative response process has three ordered response categories for each statement, the response process for the middle category is single-peaked and therefore is of the unfolding type. The model is produced by using this feature for each statement and folding over the two extreme categories. Constructing the model requires a shift of perspective from the traditional approach: In the traditional approach, the focus is on the data that must be unfolded using a model, whereas in the present approach the focus is on an existing model that is already unfolded, but that must be folded to correspond to the data.

The cumulative model used with statements that contain ordered response categories is an extension of Rasch's (1960/1980) model for dichotomous responses which, in turn, is the probabilistic counterpart of the Guttman scale (Andrich, 1985). An advantage of a formal model, in addition to summarizing and explaining the features of phenomena of interest, is that through its own logic it can add new insights into the original phenomena. Sometimes the logic of a model not only leads to new insights, but to surprising insights that otherwise would not be considered. Genuine advances are made when a conclusion or deduction is surprising in the first instance, but which on reflection appears tenable. An outcome of this kind in the model to be derived is that a unit of measurement for each statement, or for the entire scale, arises directly in an unusual way, but which, on closer examination, is shown to be consistent with the way a unit operates in typical measurement.

The Hyperbolic Cosine Unfolding Model

The Unfolding Response Process

Consider a dichotomous response of *Agree* or *Disagree* when a person encounters a statement in which the person and the statement are located on the same linear continuum. Further, consider the statement *I don't believe in capital punishment but I am not sure it isn't necessary*, which can be considered to reflect an ambivalent attitude toward capital punishment.

Let β_n be the location of person n and δ_i be the location of statement i. If the person's location

is very close to the statement location (ambivalent or moderate about capital punishment), the person will tend to agree with the statement. As the person's location increases in distance from the statement *in either direction* (either very much for or very much against capital punishment), the probability of the person selecting Agree will decrease. This gives the single-peaked form of the probability of the Agree response, which is the curve that is usually focused on exclusively.

However, now consider the complementary Disagree response. The probability of a Disagree response will increase as the person's location becomes greater than that of the statement ($\beta_n >> \delta_i$, very much in favor of capital punishment) and as the person's location becomes much less than the location of the statement ($\beta_n << \delta_i$, very much against capital punishment). Thus a person may select the Disagree response for two reasons.

Therefore, a Disagree response gives no indication of the direction in which the person is distant from the statement; this gave rise to the term "unfolding" (Coombs, 1950, pp. 147–150). That is, the continuum had both directions aligned with each other at the location of the person (the person's "ideal point"); it needed to be "unfolded" in order to establish a continuum. Thus, although the term unfolding can be related to substantive psychological processes, it arises out of the task required to construct a linear continuum in which direction has meaning, from responses that provide no information about direction.

Now consider keeping the two reasons for the Disagree response distinct. Then, the following three latent responses result in the two manifest responses of Agree and Disagree: (1) disagree because the person considers him/herself below the location of the statement [disagree below (DB)]; (2) agree because the person considers him or herself close to the location of the statement [agree close (AC)]; and (3) disagree because the person considers him/herself above the location of the statement [disagree above (DA)].

Furthermore, the probabilities of DB and DA have a monotonic decreasing and increasing structure, respectively, just as a response has in the cumulative process. Making the two reasons for a Disagree response explicit is exploited in making a cumulative model for three ordered categories relevant to the unfolding process, and is the distinctive feature in constructing the final model.

The Rasch Model for Ordered Response Categories

The Rasch model for ordered response categories has been derived and presented sufficiently often (Andersen, 1977; Andrich, 1978, 1979; Wright & Masters, 1982) that it can be stated directly. If the $m + 1$ successive ordered categories of statement i are associated with the random variable X, having values $x \in \{0, 1, 2, \ldots, m\}$, then when person n responds to statement i, the probability of the response x_{ni} takes the form

$$P\{X = x_{ni}\} = \frac{1}{\gamma_{ni}} \exp[\kappa_{xi} + x(\beta_n - \delta_i)] , \qquad (1)$$

where the κ_{xi} are the coefficients of each category x for each statement i and can be resolved according to

$$\kappa_{oi} = \kappa_{mi} \equiv 0 \qquad (2)$$

and

$$\kappa_{xi} = -\sum_{k=1}^{x} \tau_{ki} , \qquad (3)$$

where τ_{ki} are thresholds on the ordered continuum that separate the successive categories. β_n and

δ_i are the locations of person n and statement i, respectively, and

$$\gamma_{ni} = \sum_{x'=0}^{m} \exp[\kappa_{x'i} + x'(\beta_n - \delta_i)] \tag{4}$$

is a normalizing factor that ensures that the probabilities in Equation 1 sum to 1.0. The definition in Equation 2 specifies the required constraints on the thresholds. For convenience, and except when it helps the derivation of equations, the value x is not subscripted by n and i, it being understood that it is the response of person n to statement i.

The category response functions (CRFs) for a statement with three ordered categories (and two thresholds) are shown in Figure 1. This is used to derive the unfolding model. It is evident from the three CRFs that as the location β_n of person n increases relative to the location δ_i of statement i, the probability of the highest score of 2 increases, and as the location β_n decreases relative to the location δ_i, the probability of the score 0 increases.

Figure 1
CRFs for Three Response Categories: Categories Scored $x = 0$ and $x = 2$
Correspond to DB and DA, Respectively, and $x = 1$ Corresponds to AC
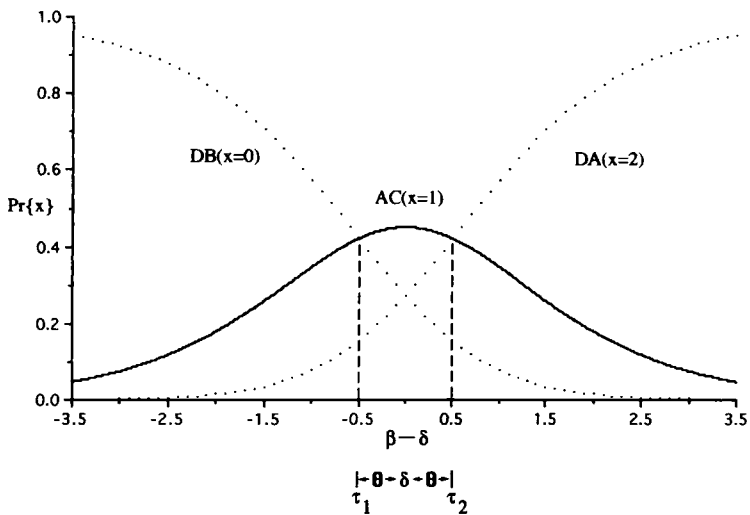
In the general case in which the thresholds between the successive categories are equal (i.e., $\tau_{x+1,i} - \tau_{xi} = \tau_{xi} - \tau_{x-1,i}$ for all x), Equation 1 may be written in the simpler form

$$P\{x_{ni}\} = \frac{1}{\gamma_{ni}} \exp[x(m - x)\theta_i + x(\beta_n - \delta_i)] \ , \tag{5}$$

where $\theta_i = (\tau_{xi} - \tau_{x-1,i})/2$, and is taken as positive (Andrich, 1982). Clearly, when a statement has more than two thresholds (and three categories), the form of Equation 5 provides a reduced parameterization of the model because only a single distance between all pairs of successive thresholds is specified rather than the value of each threshold.

However, in the case of two thresholds (and three categories), which is of interest here, the form of Equation 5 specifies completely the parameters of Equation 1 and there is no reduction in the parameterization. Note also that in the three-category case the location parameter δ_i of the statement is in the middle of the two thresholds, and θ_i represents the distance from δ_i to each of the

thresholds. With only three categories (i.e., $m = 2$), it is instructive to be explicit about the probabilities of the values of x:

$$P\{x = 0\} = \frac{1}{\gamma_{ni}} , \tag{6}$$

$$P\{x = 1\} = \frac{1}{\gamma_{ni}} \exp(\theta_i + \beta_n - \delta_i) , \tag{7}$$

and

$$P\{x = 2\} = \frac{1}{\gamma_{ni}} \exp 2(\beta_n - \delta_i) . \tag{8}$$

The threshold τ_{ki} is defined as the point on the continuum on either side of the statement location in which the probability of a response in the categories on either side of the threshold are equal. Thus, θ_i is the distance on either side of the statement location in which the probability of each extreme response becomes greater than the middle response. In a sense it is the half unit of measurement about the statement location on the latent continuum. For example, in having points on a ruler marked in centimeters, the location of any object deemed to be within half a centimeter on either side of a point will be assigned the length represented by that point. In the same way, any location of β_n within θ_i from δ_i will have the highest probability of being located in the middle category, and therefore at the location of δ_i with respect to that item. This feature is explored in greater detail throughout this paper.

The Unfolding Model With a Location and Unit for Each Statement

The unfolding response process and the cumulative response model are brought together in the following steps. First, although the model in Equations 6-8 is in a cumulative framework, it is evident from the CRFs of Figure 1 that the probability of the response in the middle category has a single peak. Thus, the response to the middle category characterizes a response of the unfolding type: Specifically, as the distance between the person's location β and the statement's location δ increases in either direction, the probability of the response in the middle category tends to 0.0.

Second, if the middle category ($x = 1$; Equation 7) represents the Agree response, the probability of AC has the required unfolding structure. That is, AC can be mapped onto $x = 1$. However, in Figure 1 ($x = 0$) represents DB and ($x = 2$) DA, and these response functions are monotonically increasing (DA) and decreasing (DB).

Interestingly, and conveniently for the next step in the derivation of the unfolding model, these CRFs are made explicit, rather than implicit, as they generally are in traditional developments for unfolding models for direct responses. Thus, the CRF structure of the three response categories of a model that is normally considered exclusively cumulative, is in fact also a model for the unfolding of dichotomous responses—because there are three, not two, implicit responses in the dichotomous unfolding response process, and they have exactly the same structure as the cumulative model for three categories. This observation leads to the construction of the new unfolding model for a direct response design.

Third, recognizing that the two reasons for a Disagree response are not distinguished in the data leads to the next step of summing or pooling the probabilities of the two categories in the model in Equations 6-8. This step may be taken as folding over the two latent Disagree responses in the model in order to correspond to the data. Thus, rather than focusing on the data and seeking devices

for unfolding them, this approach focuses on a model that assumes that data are already unfolded, and then folds the model to correspond to the data.

Consequently, the sum of the probabilities of $x = 0$ (DB) and $x = 2$ (DA) represents the probability of the one manifest Disagree response and, in the data, this Disagree response may be scored $x = 0$. Using Equations 6–8, in which Equations 6 and 7 are summed, the construction gives the probabilities of the required model:

$$P\{x = 1\} = \frac{\exp(\theta_i + \beta_n - \delta_i)}{1 + \exp(\theta_i + \beta_n - \delta_i) + \exp 2(\beta_n - \delta_i)} \tag{9}$$

and

$$P\{x = 0\} = 1 - P\{x = 1\} = \frac{1 + \exp 2(\beta_n - \delta_i)}{1 + \exp(\theta_i + \beta_n - \delta_i) + \exp 2(\beta_n - \delta_i)}, \tag{10}$$

where $x = 1$ represents AC, and $x = 0$ represents both of the (latent) reasons for the (single manifest) Disagree response (DB and DA).

Note that in this derivation another random variable could have been introduced for Equations 9 and 10; however, because the probability of $x = 1$ representing the Agree response is not affected, the same random variable X has been retained. Also, the pooling of the probabilities for the two reasons for a Disagree response is analogous to constructing the two-tail region for rejecting a statistical null hypothesis. The hypothesis is rejected if the estimated parameter location is greater or less than the critical values, and the probability of doing so is the sum of the one-tail probabilities.

Equations 9 and 10 can be simplified considerably. First, multiply the numerator and denominator by $\exp(-\beta_n + \delta_i)$, resulting in

$$P\{x = 1\} = \frac{\exp(\theta_i)}{\exp\theta_i + \exp(-\beta_n + \delta_i) + \exp(\beta_n - \delta_i)} \tag{11}$$

and

$$P\{x = 0\} = \frac{\exp(-\beta_n + \delta_i) + \exp(\beta_n - \delta_i)}{\exp\theta_i + \exp(-\beta_n + \delta_i) + \exp(\beta_n - \delta_i)}. \tag{12}$$

Second, recognizing that $[\exp(-a) + \exp(a)]/2 = \cosh(a)$, the hyperbolic cosine, gives

$$P\{x = 1\} = \frac{1}{\gamma_{ni}} \exp(\theta_i) \tag{13}$$

and

$$P\{x = 0\} = \frac{1}{\gamma_{ni}} 2\cosh(\beta_n - \delta_i), \tag{14}$$

where the normalizing factor is $\gamma_{ni} = \exp(\theta_i) + 2\cosh(\beta_n - \delta_i)$.

This model is very simple. The hyperbolic cosine, a symmetric function, provides the symmetry required to reflect the two reasons for the Disagree response. It is generically referred to, therefore, as the hyperbolic cosine model (HCM). The probabilities of the two responses, Agree and Disagree, according to Equations 13 and 14 are shown in Figure 2.

The model also provides a location (δ_i) and unit (θ_i) parameter for each statement, which are in

Figure 2
CRFs for Two Observed Responses,
$x = 0$ (Disagree) and $x = 1$ (Agree)
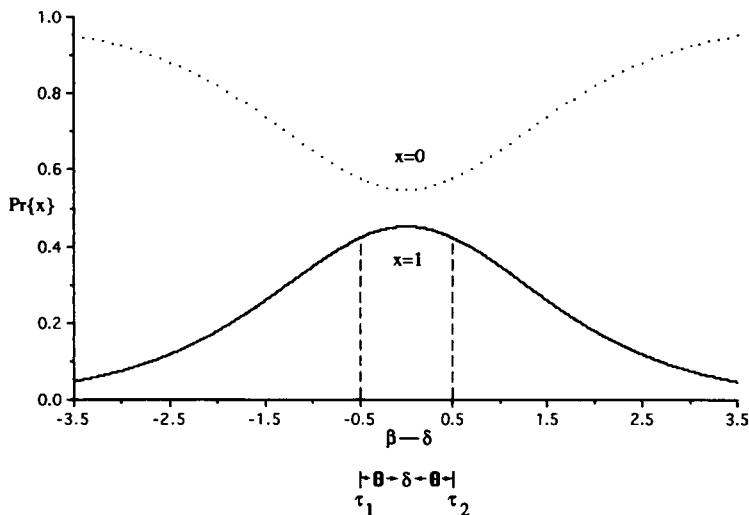
additive form. This contrasts with the multiplicative form, such as $\alpha(\beta - \delta)$, that a unit-type parameter α (usually termed a discrimination parameter) takes in other latent trait models; this, therefore, raises new issues as to the operation of a unit in this model.

For a fixed θ_i, the greater the difference $(\beta_n - \delta_i)$ in either direction, the lower the probability that the response will be $x = 1$. If $\beta_n - \delta_i = 0$, then the probability of a positive response is given by

$$P\{x_{ni} = 1\} = \exp\theta_i/(2 + \exp\theta_i) , \tag{15}$$

which indicates that the response is a function of the size of the unit parameter. Then, as the unit θ_i increases, the range on the continuum within which a person is likely to perceive the statement as being close enough to agree to the statement increases, and the probability of an Agree response increases.

When the unit parameter becomes $\theta_i = 0$ (i.e, when the thresholds on the continuum are located at identical positions), the probability of a positive response when $\beta_n = \delta_i$ is $P(x = 1)$ is .33. This indicates that each of the three latent responses—DB, AC, and DA—has the same probability of being observed. Of course, the two reasons for a negative response are not distinguished, so the total probability of a negative response is .67.

This uniform distribution for the three latent responses is consistent with the uniform distribution under symmetric cumulative models (the Rasch and the two-parameter logistic models) for dichotomous data in which the probability of each response is .5 when $\beta_n = \delta_i$. The meaning of the unit θ_i is elaborated below, but for the present note that $\theta_i = 0$ is a boundary condition and that the latent continuum is continuous; therefore, the probability of any person or any statement being located at any position is 0. In the model, θ may be less than 0. However, if $\theta_i < 0$, then the response process becomes U-shaped (Andrich, 1982): A negative estimate of θ indicates immediately that the unfolding (single-peaked) response process is not a property of the data. Figure 3 shows the response probabilities for three values of θ_i.

Figure 3
Probability of an Agree Response as a Function of the Relative Location
of the Person and the Statement ($\beta - \delta$) for Three Values of (θ)
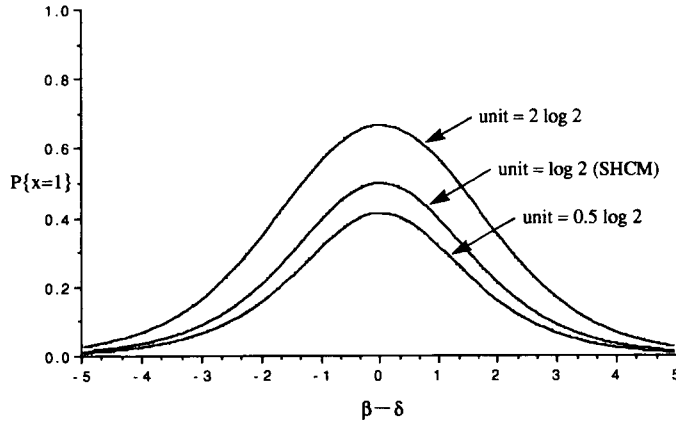
The Simple HCM—A Constant Unit

The model of Equations 13 and 14 has a different unit parameter θ, for each statement i. The model can be specialized by making $\theta_i = c$ (a constant) for all items. In some sense this specialization is equivalent to that of specifying equal discriminations in the two-parameter logistic model for cumulative dichotomous responses (which then reduces to the Rasch model) and to that of specifying equal variances in the discriminal processes in Thurstone's law of comparative judgment (which leads to his Case V; Thurstone, 1927).

Rather than setting $\theta_i = 1$, a more judicious selection may be made that provides a very elegant model. First, Equation 13 may be written explicitly in the form

$$P\{x = 1\} = \frac{\exp(\theta_i)}{\exp(\theta_i) + 2\cosh(\beta_n - \delta_i)} , \tag{16}$$

and if $\exp(\theta_i = 2)$ (i.e., $\theta_i = \ln 2$ for all i), the model simplifies to

$$P\{x = 1\} = \frac{1}{1 + \cosh(\beta_n - \delta_i)} \tag{17}$$

and

$$P\{x = 0\} = \frac{\cosh(\beta_n - \delta_i)}{1 + \cosh(\beta_n - \delta_i)} , \tag{18}$$

or in terms of a single expression

$$P\{x\} = \frac{\cosh^{1-x}(\beta_n - \delta_i)}{1 + \cosh(\beta_n - \delta_i)} . \tag{19}$$

This model may be seen as the unfolding counterpart to the Rasch model for cumulative dichotomous responses, in that $\cosh(\beta_n - \delta_i)$ is made up of (is the average of) the complementary elements of the Rasch model [i.e., $\exp(\beta_n - \delta_i)$ and $\exp(\delta_i - \beta_n)$]. However, the model of Equations 16–19 does

not belong to the Rasch class of models because it has no sufficient statistics. In addition, the selection of the specific value of $\theta_i = \ln 2$ is made for mathematical convenience.

The model of Equations 16–19, which contains only the parameters β_n and δ_i in the form $(\beta_n - \delta_i)$, is parametrically equivalent to the squared simple logistic model (SSLM) introduced by Andrich (1988), which also contains the same parameters in the same form $(\beta_n - \delta_i)$. In the SSLM, the square of $(\beta_n - \delta_i)$ is taken in order to create the symmetry required to produce a single-peaked response process; but in the HCM, the symmetric hyperbolic cosine of $(\beta_n - \delta_i)$ is taken. This latter model will be termed the simple HCM (SHCM), in parallel to the SSLM. The rationale for the SHCM—which involves analyzing the details of the unfolding response mechanism to reveal and then model explicitly the two latent responses for a Disagree response, and then bringing them together—is more compelling than that for the SSLM, which is essentially based on a device (the square function) that produces a descriptive model of the required shape.

One of the properties of the SSLM is that the maximum probability for a response, when $\beta_n - \delta_i = 0$, is only .5. Remarkably, that is also the case for the SHCM. Hoijtink (1990) has generalized the SSLM to permit greater probabilities. Greater probabilities with the general HCM are possible by retaining the unit parameter θ_i, as in Equations 13 and 14. The model that incorporates the unit parameter θ_i for each item will be termed the two-parameter HCM (2PHCM). Although the emphasis here is on the SHCM, there is also a discussion, development, and application of the 2PHCM.

Maximum Likelihood Estimates

The literature on latent trait models suggests that a number of procedures for estimating parameters may be used. However, the procedures specific to the Rasch class of models—those characterized by the presence of sufficient statistics—are not among them. This is because the structure of the Rasch models is destroyed by summing or pooling two of the categories in Equations 6–8.

The method used is termed the joint maximum likelihood (JML) estimation procedure. JML is based on the joint likelihood of the data with respect to both persons and statements and was a very common method used in latent trait scaling until the advent of the EM algorithm and marginal maximum likelihood methods. In addition to being simple, the solution equations for the JML procedure are instructive in understanding some features of the unfolding process. Furthermore, no assumptions of any distribution of person parameters need to be made.

From the work on estimation in the Rasch models (Andersen, 1973; Jansen, Van den Wollenberg, & Wierda, 1988; Wright, 1988; Wright & Douglas, 1977), it can be anticipated that for a fixed number of statements and with an increasing number of persons, the estimates of the statement parameters will not be consistent, but that this inconsistency should be manageable. This point is elaborated below.

For the random variable x_{ni} that takes on the value 1 for a positive response (Agree) and 0 for a negative response (Disagree), the probabilities in Equations 9 and 10 can be expressed in the form

$$P\{x_{ni}\} = \frac{1}{\gamma_{ni}} \cosh^{1-x_{ni}}(\beta_n - \delta_i) \ . \tag{20}$$

The joint likelihood of the matrix of responses $\mathbf{X}$ of dimension $N \times I$, given by

$$L = \prod_n \prod_i P\{x_{ni}\} \ , \tag{21}$$

provides the log-likelihood equation

$$\mathrm{Log}\, L = \sum_i \sum_n [(1 - x_{ni}) \log \cosh(\beta_n - \delta_i)] - \sum_i \sum_n \ln \gamma_{ni} \ . \tag{22}$$

Clearly, two sets of equations are required—one for the δ_i, the other for the β_n. In forming these, let π_{ni} denote the probability $P(x_{ni} = 1)$. Differentiating $\log L$ with respect to δ_i, setting the derivative equal to 0, and simplifying gives the first set of equations:

$$
\begin{aligned}
\phi(\delta_i) = \frac{\partial \log L}{\partial \delta_i} &= \sum_n (1 - x_{ni}) \frac{\sinh(\beta_n - \delta_i)}{\cosh(\beta_n - \delta_i)} - \sum_n \frac{2\sinh(\beta_n - \delta_i)}{\gamma_{ni}} \\
&= \sum_n (1 - x_{ni}) \tanh(\beta_n - \delta_i) - \sum_n \frac{2\cosh(\beta_n - \delta_i)\tanh(\beta_n - \delta_i)}{\gamma_{ni}} \\
&= \sum_n (1 - x_{ni}) \tanh(\beta_n - \delta_i) - (1 - \pi_{ni}) \tanh(\beta_n - \delta_i) \\
&= \sum_n (\pi_{ni} - x_{ni}) \tanh(\beta_n - \delta_i) \ .
\end{aligned}
\tag{23}
$$

Therefore, the solution equation is

$$
\phi(\delta_i) = \sum_n (\pi_{ni} - x_{ni}) \tanh(\beta_n - \delta_i) = 0 \ .
\tag{24}
$$

Analogously, the solution equation for the second set β_n is given by

$$
\phi(\beta_n) = \sum_i (\pi_{ni} - x_{ni}) \tanh(\beta_n - \delta_i) = 0 \ .
\tag{25}
$$

These equations are structurally analogous to those of the SSLM.

In this system of equations in which the sum over n for Equation 25 is the same as the sum over i for Equation 24, one constraint is required. The usual kind of constraint is to set

$$
\sum_i \hat{\delta}_i = 0 \ .
\tag{26}
$$

No other constraint is necessary.

These equations are related to those for the cumulative Rasch model for dichotomous responses. In particular, Equations 24 and 25 contain the terms

$$
\sum_n (x_{ni} - \pi_{ni})
\tag{27}
$$

and

$$
\sum_i (x_{ni} - \pi_{ni}) \ ,
\tag{28}
$$

respectively, and these are augmented by the factor $\tanh(\beta_n - \delta_i)$ that characterizes the unfolding aspect of the data through the expression $(\beta_n - \delta_i)$.

The simultaneous solution of Equations 24 and 25 must be obtained iteratively. The Newton-Raphson algorithm, which requires using the second derivatives, is considered here. These second derivatives, however, are instructive in themselves; their derivation is summarized in the Appendix. In addition, these second derivatives contain data. Therefore, for the parameter estimation and for the subsequent estimation of the SEs of the estimates, their expectations must be taken. These expectations give the Fisher information functions, and take the forms

$$
E\left[\frac{\partial^2 \log L}{\partial \delta_i^2}\right] = -\sum_n \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i)
\tag{29}
$$

and

$$E\left[\frac{\partial^2 \log L}{\partial \beta_n^2}\right] = -\sum_i \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) \ . \tag{30}$$

A Solution Algorithm

A direct application of the multiparameter Newton-Raphson algorithm would take the form

$$
\begin{matrix}(I-1) \\ (N)\end{matrix}
\begin{bmatrix}\hat{\delta} \\ \hat{\beta}\end{bmatrix}_{(q+1)}
=
\begin{bmatrix}\hat{\delta} \\ \hat{\beta}\end{bmatrix}_{(q)}
-
\begin{bmatrix}\psi_{\delta\delta} & \psi_{\delta\beta} \\ \psi_{\beta\delta} & \psi_{\beta\beta}\end{bmatrix}
\begin{bmatrix}\hat{\delta} \\ \hat{\beta}\end{bmatrix}_{(q)} , \tag{31}
$$

where

$$\psi_{\delta\delta} = E\left[\frac{\partial^2 \log L}{\partial \delta_i \partial \delta_i}\right] , \tag{32}$$

$$\psi_{\beta\beta} = E\left[\frac{\partial^2 \log L}{\partial \beta_n \partial \beta_n}\right] , \tag{33}$$

and

$$\psi_{\beta\beta} = E\left[\frac{\partial^2 \log L}{\partial \delta_i \partial \beta_n}\right] , \tag{34}$$

are the submatrices of the matrix of second derivatives and $[d:b]'_{(q)}$ is the vector of provisional estimates at the qth iteration.

This multiparameter algorithm is not feasible because as N increases, the matrix of second derivatives becomes too large to invert. In the alternative algorithm, used in the program for the analyses reported in this paper, the provisional statement parameters are regarded as fixed, and the Newton-Raphson algorithm is employed to improve the estimates of the person parameters according to

$$[\hat{\beta}_n]_{(q+1)} = [\hat{\beta}_n]_{(q)} - [\psi_{\beta\beta}]_{(q)}^{-1}\phi(\hat{\beta}_n)_{(q)} , \tag{35}$$

where the submatrix $\psi_{\beta\beta}$ is readily inverted because it is simply diagonal. With the provisional person parameters then taken as fixed, the statement location parameters δ_i are estimated according to

$$[\hat{\delta}_i]_{(q+1)} = [\hat{\delta}_i]_{(q)} - [\psi_{\delta\delta}]_{(q)}^{-1}\phi(\hat{\delta}_i)_{(q)} . \tag{36}$$

Following the improved estimates of δ to a specified degree of accuracy, the restriction

$$\sum_i \hat{\delta}_i = 0 \tag{37}$$

is imposed and the entire process is repeated until all the item parameters as a set converge to a selected level. Incidentally, finite estimates for persons and statements can be obtained if all responses are 1s. Only if all responses are 0s can no finite estimate be obtained. These estimates will be the same as if the full matrix containing the off-diagonal terms in Equation 31 were inverted in the algorithm, and only the rate of convergence and the estimates of the SEs are affected.

Initial Estimates

A judicious selection of initial estimates is helpful in obtaining rapid convergence. The initial

location values $d_i^{(0)}$ are obtained by setting $\beta_n = 0$ for all n, giving

$$\frac{s_i}{N} = \frac{1}{\cosh \hat{\delta}_i^{(0)} + 1} \ , \tag{38}$$

where

$$s_i = \sum_n x_{ni} \tag{39}$$

is the total score of a statement, and

$$\hat{\delta}_i^{(0)} = \cosh^{-1}\left(\frac{N - s_i}{s_i}\right) . \tag{40}$$

However, because $\cosh\delta \geq 1$, it is necessary that $(N - s_i)/s_i \geq 1$ for an initial estimate according to Equation 40, which may not be satisfied in some datasets.

There are different ways to ensure that a value is obtained, and the one used in the program developed for the analyses in this paper is as follows. If for any i, $(N - s_i)/s_i < 1$, then the minimum of these values is taken, the difference

$$\alpha = 1 - \min_i \ (N - s_i)/s_i \tag{41}$$

is calculated, and the initial estimates are formed according to

$$\hat{\delta}_i^{(0)} = \cosh^{-1}[\alpha + (N - s_i)/s_i] . \tag{42}$$

The unfolding nature of the data creates one further complication: Statements at genuinely different locations can have the same statistic (total score) s_i. This difference is accommodated by incorporating both negative and positive values into the initial estimates, giving

$$\hat{\delta}_i^{(0)} = \pm\cosh^{-1}[\alpha + (N - s_i)/s_i] . \tag{43}$$

In general, the provisional ordering of the statements—to the degree that some are given negative signs and some are given positive signs around the constraint that

$$\sum_i \hat{\delta}_i = 0 \tag{44}$$

—can be hypothesized from theoretical considerations. However, for the more empirically oriented, even though it does not arise from the theory developed here, the sign can be obtained from the signs of the statement loadings on the first factor in a factor analysis. As shown in Davison (1977), if the data exhibit an unfolding structure, then these loadings exhibit a corresponding structure in which the signs reflect the direction of the statement.

In obtaining the initial location values β_n, advantage is taken of the initial values $\delta_i^{(0)}$:

$$\hat{\beta}_n^{(0)} = \sum_{i=1}^{I_n} x_{ni}\hat{\delta}_i^{(0)}/I_n \ , \tag{45}$$

where I_n is the number of items that person n has with a score of 1. These values are the average of the location values of the statements agreed to by the person.

SEs of Estimates and Information

According to maximum likelihood theory, variances of the estimates of β_n and δ_i are approximated

according to

$$\hat{\sigma}^2_{\beta_n} = -E\left[\frac{\partial^2 \log L}{\partial \beta_n^2}\right] = 1 \bigg/ \sum_i \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) \tag{46}$$

and

$$\hat{\sigma}^2_{\delta_i} = -E\left[\frac{\partial^2 \log L}{\partial \delta_i^2}\right] = 1 \bigg/ \sum_n \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) \ , \tag{47}$$

respectively. Of course, using these expressions includes the approximation in which the off-diagonal elements of Equation 31 are ignored. Strictly, the diagonals of the inverted matrix of Equation 31 should be used rather than simply the inversions of the diagonals of the submatrices of Equations 29 and 30. Nevertheless, these terms may be used as approximations and to gain insight into the unfolding response mechanism. The effect of not using the off-diagonals is noted in the simulation study reported below, and is shown to be relatively small.

Consider the case of a single person responding to a single statement, and denote the information I with respect to $\beta - \delta$ associated with the response as

$$I = I_{\beta\delta} = \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) \ . \tag{48}$$

Because information with respect to a person or statement parameter is additive, the effect of having many persons or many statements can be determined immediately. As with the SSLM developed in Andrich (1988), $I = 0$ for $\beta = \delta$, which contrasts with symmetric cumulative response models in which maximum information is obtained when $\beta = \delta$. Again, as with the SSLM, the information function is bimodal, reaching a maximum at a value of $|\beta_n - \delta_i| > 0$ for a given θ_i, and then having an asymptote of 0 as $|\beta_n - \delta_i| \to \infty$. Figure 4 shows this information function with respect to $\beta_n - \delta_i$ for the SHCM. A full discussion of the interpretation of a bimodal information function for unfolding models and responses is provided in Andrich (1988).

Figure 4
Information in a Single Response as a Function of the Relative
Location of the Person and the Statement ($\beta - \delta$) for the SHCM
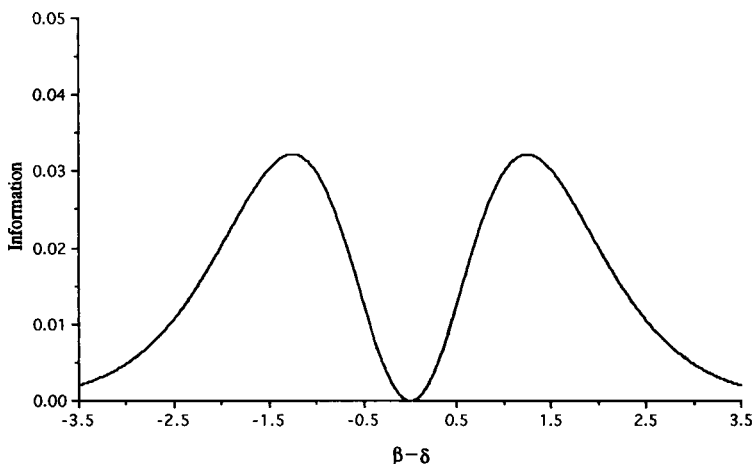
Inconsistency of Parameter Estimates

It has already been indicated that the estimates of the parameters are not expected to be consistent. Although concern here is mostly with the introduction of the model and demonstrating that it is feasible to apply, the issue of consistency is broached for completeness. First, with 15 to 20 statements and 500 persons, the inconsistency seems noticeable. On the other hand, when the number of statements is 100, the effect of the inconsistency seems minute. From this, it is inferred that as the number of statements and persons becomes very large, the estimates will converge to their expected values. This result, yet to be proven, is analogous to that established by Haberman (1977) for the JML estimates in the simple Rasch model for dichotomous responses.

Estimation of the Unit Parameters

If the unit parameter is permitted to vary, as in the 2PHCM (Equations 13, 14, and 16), and this parameter is to be estimated, then the solution equations need to be augmented by

$$\phi(\theta_i) = \sum_i (\pi_{ni} - x_{ni}) = 0 \tag{49}$$

and

$$E\left[\frac{\partial^2 L}{\partial \theta_i^2}\right] = \sum_n \pi_{ni}(i - \pi_{ni}) \ . \tag{50}$$

In the 2PHCM, however, the estimates of the unit and location parameters are correlated. This is not unusual—the estimates of the difficulty and discrimination parameters in the two-parameter logistic model for cumulative responses also are correlated. However, this correlation may have a special significance in the case of mapping locations on a continuum, and so these effects are broached in both the simulation study and the analysis of real data. This correlation would be evident in any method of estimation (cf. Verhelst & Verstralen, 1991) because it is not a property of the estimation but of the model.

Example Applications

Simulation Study

The simulation involved the SHCM and 20 statements. This is a relatively large number of statements for unfolding data, but is quite typical of the number of statements in attitude measurement. The statement locations were spread uniformly (but in pairs) across the attitude continuum in the range –3.15 to 3.15. The simulation was repeated five times with 500 simulated persons in each replication. The parameters were drawn from a normal distribution with mean $\mu = 0.0$ and a standard deviation $\sigma = 2.5$. A convergence criterion of .01 was used for the increments in the solution of the implicit equations.

Recovery of statement parameters.   Table 1 shows the generating parameters of the statements, their estimate in each replication, and the average of the estimates across replications. It also shows the correlation between the estimates and the generating parameters. The recovery of the statement location parameters was excellent. Moreover, the standard deviations (SDs) of the estimates and the SEs obtained from Equation 47 were relatively similar. The average of the estimates of the SEs (ASE) for all items according to Equation 47 was .158 and the average of the SDs of estimates across the five replications over all items was .152. Thus, taking only the diagonal elements of the matrices of

Table 1
Generating Parameters (GP), Estimates of Item Location in Five Replications for the
SHCM, Their Variances (Var), Standard Deviations (SD), and Average Standard
Errors (ASE), and Correlations (r) of GP and the Estimates; and Estimates of
the Unit Parameters Under the 2PHCM Where the Unit = log 2.0 = .693

Item	GP	Replication 1	2	3	4	5	Mean	SD	ASE
SHCM: Locations									
1	−3.150	−3.418	−3.025	−3.217	−3.323	−3.342	−3.265	.152	.176
2	−3.150	−3.341	−3.368	−3.356	−3.336	−3.409	−3.362	.029	.179
3	−2.450	−2.661	−2.450	−2.724	−2.595	−2.507	−2.587	.111	.163
4	−2.450	−2.542	−2.535	−2.398	−2.751	−2.842	−2.614	.179	.163
5	−1.750	−1.959	−1.897	−2.139	−2.156	−1.842	−1.999	.142	.155
6	−1.750	−1.993	−1.823	−1.917	−1.989	−1.620	−1.868	.155	.154
7	−1.050	−.985	−1.271	−1.018	−1.154	−1.225	−1.131	.126	.150
8	−1.050	−1.117	−1.086	−1.261	−.733	−1.018	−1.043	.195	.149
9	−.350	−.395	−.507	−.130	−.645	−.468	−.429	.190	.148
10	−.350	−.355	−.867	−.322	−.329	−.770	−.528	.267	.148
11	.350	.614	.665	.269	.477	.271	.459	.186	.148
12	.350	.523	.272	.453	.375	.437	.412	.095	.148
13	1.050	1.055	1.260	1.159	.930	.887	1.058	.155	.149
14	1.050	1.134	1.129	1.111	1.050	1.184	1.122	.048	.150
15	1.750	1.896	1.774	1.672	1.968	2.026	1.867	.144	.154
16	1.750	1.830	1.702	1.744	1.694	1.791	1.752	.058	.153
17	2.450	2.274	2.545	2.672	2.858	2.898	2.649	.254	.162
18	2.450	2.405	2.817	2.630	2.863	2.759	2.695	.184	.162
19	3.150	3.335	3.371	3.610	3.396	3.179	3.378	.155	.176
20	3.150	3.701	3.295	3.162	3.400	3.610	3.433	.222	.177
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.152	.158
Var	4.256	4.866	4.722	4.818	5.054	5.022	4.875	.004	.0001
SD	2.063	2.206	2.173	2.195	2.248	2.241	2.208	.063	.011
r		.998	.997	.998	.997	.997	.999		
2PHCM: Units									
1	.693	.745	.432	.629	.802	.927	.707	.188	.145
2	.693	.754	1.036	.673	.463	.311	.648	.278	.148
3	.693	.833	.653	.765	.591	.792	.727	.101	.130
4	.693	.454	.722	.913	.592	.787	.694	.177	.131
5	.693	.660	.667	.797	.666	.914	.741	.112	.121
6	.693	.646	.561	.798	.741	.657	.681	.091	.121
7	.693	.801	.685	.636	.603	.861	.718	.110	.113
8	.693	.598	.581	.501	.753	.785	.644	.121	.114
9	.693	.592	.575	.790	.578	.780	.663	.112	.111
10	.693	.580	.518	.754	.763	.787	.680	.122	.111
11	.693	.752	.650	.689	.784	.651	.705	.060	.110
12	.693	.679	.731	.691	.754	.452	.661	.121	.111
13	.693	.869	.548	.803	.599	.816	.727	.144	.113
14	.693	.709	.648	.525	.877	.711	.694	.127	.113
15	.693	.723	.730	.772	.635	.713	.715	.050	.119
16	.693	.625	.716	.661	.655	.677	.667	.033	.119
17	.693	.838	1.038	.463	.678	.418	.687	.259	.130
18	.693	.599	.956	.647	.903	.432	.708	.219	.131
19	.693	.717	.868	.836	.568	.500	.698	.162	.146

continued on the next page

Table 1, continued
Generating Parameters (GP), Estimates of Item Location in Five Replications for the
SHCM, Their Variances (Var), Standard Deviations (SD), and Average Standard
Errors (ASE), and Correlations (r) of GP and the Estimates; and Estimates of
the Unit Parameters Under the 2PHCM Where the Unit = log 2.0 = .693

Item	GP	Replication					Mean	SD	ASE
		1	2	3	4	5			
20	.693	.689	.544	.517	.855	.890	.699	.172	.147
Mean		.693	.693	.693	.693	.693	.693	.1380	.1241
Var		.011	.028	.015	.014	.032	.001	.0041	.0002
SD		.103	.168	.123	.118	.180	.027	.0641	.0134

second derivatives and excluding the off-diagonal elements had a negligible effect on the SEs.

To examine the effect of the inconsistency of the estimates of location values, the variance of the estimates of these locations in each replication was compared to the variance of their generating values, which was 4.256. This comparison was simplified by taking the average variance of estimates across the replications, which was 4.875. Clearly 4.875 is greater than 4.256; however, this variance contains the error variance of the estimates, and on the average, its value was .025.

Therefore, the variance of the estimates without the error variance is given by 4.875 − .025 = 4.850; this value was greater than 4.256. In terms of the units in the original metric—SDs—the comparable values were $(4.850)^{1/2} = 2.20$ and $(4.256)^{1/2} = 2.06$, showing that the SDs of the estimates were greater than that of the original parameters by approximately 6%. This effect was very small, and for an analysis of a single dataset the systematic effect was effectively only a matter of scale, because only the relative parameter estimates (statement locations) were estimated and not the absolute values of a population parameter.

Table 1 also shows analogous information for the unit parameters for the 2PHCM estimation. Recall that all items were simulated with the same unit parameter in the SHCM—ln2 = .6932. Again, the estimates and their SEs were very close. In this case, if the estimated error variance $(.138^2 = .019)$ was subtracted from the average variance of the estimates (.02), the result was virtually 0 (as required because all values of the unit were equal in all replications for all items.)

The estimates of the location and unit parameters were not independent in the model. Figure 5 shows the plot of the unit estimates against the location estimates for each item in each study. As expected from the model and shown in the last column of Table 1, the spread of the estimates was greater for the extreme values of the locations. In addition, the effect of any correlation was negligible. The actual values for the correlations of each study were .086, .264, −.260, .337, and −.282 (average = .029), none of which were significant, even at the .10 level of significance.

Recovery of person parameters.   Table 2 shows a summary of the recovery of the person parameters. (In the estimation of these parameters, only those persons whose scores are all 0 cannot be estimated, which never occurred in any of the simulations). Because there were so many persons, only the variances were considered. It is evident that the estimated variance is larger than the generating variance (on the average, 5.645 compared to 4.665). If the average of the estimated error variance (.653) is subtracted from 5.645, it gives 5.645 − .653 = 4.992, which is close to the original generating variance (average of 4.665). The difference of .327 reflects the inconsistency already mentioned, but in percentage of variance terms it is only 7%, the equivalent of the effect on the statement parameters. The correlations between the generating and estimated parameters averaged .899, which was virtually the maximum value possible. This maximum was calculated by denoting $\hat{\beta}_n = \beta_n + e_n$, where

Figure 5
Estimates of the Unit Parameters and Estimates of the Location
Parameters for Statements in Five Replications for the SHCM

Replications: ○ = 1, + = 2, ● = 3, ◆ = 4, ✕ = 5.
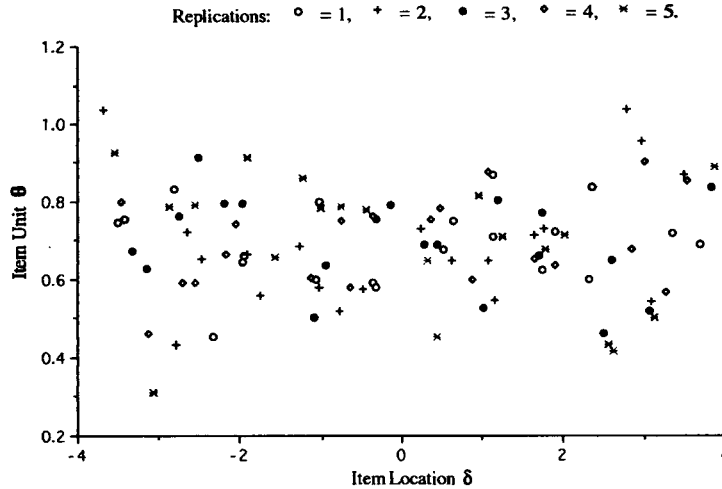
$\hat{\beta}_n$ is the estimate of the person parameter, β_n is the theoretical value, and e_n is the error. With the actual error independent of the location,

$$V[\hat{\beta}_n] = V[\beta] + V[e] . \tag{51}$$

$V[\hat{\beta}]$ can be estimated in the usual way as

$$V[\hat{\beta}] = \sum_{n=1}^{N} (\hat{\beta}_n - \hat{\beta})^2 / N - 1 , \tag{52}$$

Table 2
Generated and Estimated Person Parameters in Five Replications
for the SHCM, Their Variance (Var), Standard Deviation (SD),
Mean Squared Error (MSE), Correlation With the Generating
Parameters (r), Reliability (r_β), and Square Root of Reliability
$[(r_\beta)^{1/2} = \text{Maximum Correlation}]$

	Replication					
Statistic	1	2	3	4	5	Mean
Mean						
Generated	−.022	.026	.072	.108	−.040	.030
Estimated	−.034	.073	.117	.026	.088	.054
SD						
Generated	2.072	2.212	2.058	2.229	2.325	2.178
Estimated	2.264	2.323	2.256	2.477	2.545	2.370
Var						
Generated	4.293	4.893	4.235	4.968	5.406	4.665
Estimated	5.126	5.396	5.090	6.136	6.477	5.645
MSE	.641	.643	.643	.658	.679	.653
r	.907	.878	.883	.917	.895	.899
r_β	.875	.881	.874	.884	.895	.882
$(r_\beta)^{1/2}$	.935	.939	.935	.940	.946	.939

and $V[e]$ can be obtained as the average of the error variance for each person provided by Equation 46. Then the traditional reliability index, denoted r_β, can be estimated as

$$r_\beta = (V[\hat{\beta}] - V[e])/V[\hat{\beta}] \ . \tag{53}$$

However, from traditional test theory, the maximum correlation of a variable with error is the square root of its reliability, in this case estimated as $(r_\beta)^{1/2}$. Both r_β and $(r_\beta)^{1/2}$ are shown in Table 2.

Real Data

A small dataset presented in Andrich (1988) based on 54 graduate students replying to eight statements that were used to measure attitudes toward capital punishment was reanalyzed. The students simply agreed or disagreed to the statements, and no restriction was placed on the number of statements that were to receive an Agree response. Andrich (1988) obtained the estimates using the SSLM, and the estimates were compared with two previous scalings of these statements when data were collected according to other designs and analyzed using other models, including one from Thurstone (Wohlwill, 1963). The results from the SSLM were consistent with those obtained by the other methods; therefore, in the present analysis comparisons were made only with the results from the SSLM. In this dataset the distribution of the locations of the persons was not normal, and so an assumption of normality for this distribution would have been inappropriate in any estimation procedure. Table 3 shows the statements and the location estimates according to the SSLM and the SHCM.

In making comparisons between results, two further points need to be noted. First, the natural units in the SSLM and SHCM are not the same, but no change was made to the values obtained. Second, the 2PHCM also was used to analyze the data; the estimates of the locations and the units were then free to vary among statements. Under this model, the results shown in Table 3 indicate that there was a relatively slight effect on the location estimates. This negligible effect was shown in the correlations between the SSLM and each of the SHCM and 2PHCM where the respective values were $r = .999$ and $r = .997$; in addition, the locations of Statements 1 and 2 changed order. These statements were close in location, and the change of values was within the SE of the estimates. Thus, in this case

Table 3
Estimates of Parameters of Statements Using the SSLM, SHCM, and
2PHCM, for Statements Ordered According to the SSLM and SHCM

Statement	SSLM	SHCM	2PHCM Location	2PHCM Unit
Capital punishment is one of the most hideous practices of our time.	-.895	-1.580	-1.519	.550
The state cannot teach the sacredness of human life by destroying it.	-.845	-1.505	-1.532	.984
Capital punishment is not an effective deterrent to crime.	-.786	-1.460	-1.486	1.016
I don't believe in capital punishment but I am not sure it isn't necessary.	-.281	-.525	-.506	.535
I think capital punishment is necessary but I wish it were not.	.540	.969	1.017	.724
Until we find a more civilized way to prevent crime we must have capital punishment.	.667	1.197	1.235	.694
Capital punishment is justified because it does act as a deterrent to crime.	.880	1.411	1.343	.509
Capital punishment gives the criminal what he deserves.	.786	1.494	1.450	.534
Correlation (With SSLM)	1.00	.999	.997	

the effect on the location estimates in estimating the units was negligible.

The effect on the estimates of the person parameters is shown in Table 4, which shows the distribution of attitudes and the values obtained by the SSLM, the SHCM, and the 2PHCM, with the locations of persons and statements ordered according to the SHCM; the last column shows the rank ordering of the patterns and statements under the 2PHCM. It is evident that there were slight differences in the location values of patterns under the SHCM and the 2PHCM, including some reversals in location, which in any case were close together. Thus the effect of estimating different unit parameters for different items was very small in this example.

Table 4
Distribution of Attitudes and Their Locations Under the SSLM, the SHCM, and the 2PHCM
(Items and Patterns are Ordered Under the SHCM), and Order of Items Under the 2PHCM

SHCM Order	Response Pattern	Total Score	SSLM		SHCM		2PHCM		Fre-quency	2PHCM Order
			β	SE	β	SE	β	SE		
1	01100000	2	-1.37	.89	-2.25	1.22	-2.32	1.22	4	1
2	11100000	3	-1.17	.89	-1.92	1.33	-1.91	1.37	10	2
3	01110000	3	-.92	.82	-1.45	1.36	-1.44	1.39	3	4
4	11110000	4	-.90	.82	-1.51	1.38	-1.48	1.40	8	3
5	01111000	4	-.48	.71	-.84	1.08	-.78	1.07	1	5
6	01110010	4	-.40	.71	-.82	1.07	-.77	1.06	1	6
7	11111100	6	-.36	.70	-.71	1.02	-.65	1.01	2	7
8	10111100	5	-.21	.70	-.38	.94	-.30	.93	1	8
9	01101010	4	-.15	.70	-.22	.92	-.13	.91	1	9
10	01111110	6	0.00	.71	.03	.91	.09	.91	1	10
11	10011011	5	.22	.73	.50	1.00	.55	1.01	1	12
12	00111101	5	.23	.72	.45	.98	.52	.99	2	11
13	01011111	6	.31	.74	.74	1.11	.77	1.11	1	13
14	01001110	4	.35	.74	.86	1.18	.92	1.20	2	14
15	00101101	4	.37	.75	.89	1.20	.97	1.23	2	15
16	01001101	4	.38	.74	.89	1.20	.97	1.23	1	16
17	10001111	5	.49	.76	1.09	1.33	1.11	1.32	2	17
18	00011111	5	.64	.79	1.11	1.34	1.13	1.34	3	18
19	00010011	3	.64	.80	1.22	1.40	1.15	1.35	1	19
20	00001111	4	.95	.88	1.50	1.45	1.51	1.43	5	20
21	00001100	2	1.09	.91	1.80	1.34	1.92	1.26	1	22
22	00000111	3	1.11	.91	1.81	1.33	1.78	1.32	1	21
Mean			-.29	.78	-.44	1.27	-.42	1.27		
SD			.82	.07	1.40	.15	1.42	.16		
r with SSLM					.9945		.9940			

Summary and Discussion

The idea that θ, provides a unit of measurement has been carried through the development and application of the SHCM and the 2PHCM. According to the relationship between the data and the model, a value of the unit will be greater if the number of endorsed responses is greater. Because it may add some important new thoughts to the issue of a unit of measurement in a latent trait model, the manner in which this can be understood deserves some elaboration. In doing so, an example with an unfolding response mechanism in which the units are explicit will be considered.

Consider the case of asking a person to agree or disagree to a series of locations on a ruler in which an Agree response indicates that the particular location corresponds to the person's height. The response here will be of the unfolding kind in the sense that the person will disagree to locations

considered below his or her height and also disagree to locations above his or her height, and agree to the location considered to be at the person's height. Suppose, further, that the ruler is presented in some way so that he/she has to estimate which location to agree to with the consequence that errors are introduced.

In considering what might happen, it is necessary to be aware of three aspects of this situation. First, height is a continuous variable; therefore, the probability of the person being at any particular height is 0.0. Second, the markings designating heights on the ruler are made in units; that is, there is distance between the marks on the ruler. Third, when the person agrees to a particular marking as representing his or her height, the person is not simply agreeing to being at that mark (a point)—the person is agreeing to being within the region plus or minus one half-unit on either side of the mark. The probability of a person being located within a region has a finite nonzero number, even though the probability of being at any point is 0.

To be explicit about the implications for the unit in the unfolding process, suppose a person is 170 cm tall when measured to the nearest centimeter, and further suppose that this person is asked to agree or disagree to markings on a ruler as in the example above. Suppose further that the only marks on the ruler are 1 cm apart. The person will agree to a marking that he or she considers to be plus or minus .5 cm from his or her height. Presumably, most of the time on repeated presentations of the ruler, the person will respond to the position of 170 cm, and sometimes to 171 and sometimes to 169. The person would be expected to agree less often to the markings of 172 and 168 cm and even less often to 173 and 167 cm. Still remembering that the person is in fact agreeing to being in a region, and not at a point, suppose also that the person never agrees to markings that are 175 cm or greater, or 165 cm or lower, so that the distribution has most responses at 170 cm, and least at 174 and 166 cm. Clearly, in this case there is a distribution of responses: the person has not responded consistently.

Now suppose that instead of the markings being 1 cm apart, they are 10 cm apart with no markings in between. The person who is 170 cm tall will agree to a mark if he or she considers him or herself to be .5 of 10 cm on either side of the mark, that is to within 5 cm on either side of the mark. In centimeters, this indicates agreement to a region from 165 cm to 175 cm. If the person's perception was equally good in the two cases, then in the second case the person would always agree when presented with the mark at 170 cm, and disagree to the markings of 160 cm or less and 180 cm or more. Thus in the second case, unlike the first case, there is no variation in the responses.

This example shows that when the units are larger, the observed Agree responses are more consistent than when the units are smaller, simply because a large unit does not manifest finer distinctions in the observations. The above situation has been placed in the context of an unfolding response process, but it is, of course, no more than recognizing that if 100 persons were asked to measure the height of a person to the nearest 10 cm, they would have a greater agreement than if they were asked to measure the same person's height to the nearest 1 cm. However, even though the observed responses are more consistent (in greater agreement) when the unit of measurement is larger, the measurement itself is still less accurate than when the unit is smaller. Even if there is disagreement when measurements are taken to the nearest 1 cm and no disagreement when they are taken to the nearest 10 cm, the former is more accurate, in this case much more accurate.

In the example above, the case of a known measurement was considered, and the effect on the data when different units were taken was described. Consider now the process in reverse—the location of a person from a series of agreements or disagreements to markings on a ruler is being estimated, but the locations and units are not known. That is, consider that the only data consists of a series of Agree and Disagree responses to a set of markings on a ruler. All things being equal,

it could be inferred that in the case in which the responses are more consistent, the unit is greater. That is, the region around the location of the marking that will draw an Agree response is greater when the responses are more consistent. This is exactly the case with the HCM: The greater the number of Agree responses to a statement, all other factors being constant, the greater the estimated value of the unit θ_i.

This analysis, and the model, demonstrate that the construction of a unit from the responses is consistent with the idea that the statements have a location, but with a region around them that is like a unit of measurement. Recall that when the three CRFs—two for the two reasons for a Disagree response and one for the Agree response—are presented in full with the two tails of the Disagree response as shown in Figure 1, then θ_i represents the half distance between two thresholds of a statement, and that the location of the statement δ_i is in the middle of the two thresholds. Furthermore, these thresholds are points on the continuum around the location δ_i in which the probability of an Agree response and a Disagree response (in one direction) are equal. This provides the natural points for the region of a unit around the location δ_i in the sense that if a person considers him/herself located within the region of plus or minus θ_i from δ_i, then the person is more likely to give an Agree response than a (one-tail) Disagree response. Thus, although θ_i has been referred to as a unit, if the analogy to the measurement of length is taken literally, then it is equivalent to a half-unit. The Agree response of a person to a statement is made to the nearest $2\theta_i$ to the location δ_i of the statement.

Implications

The implications of the model for the study of a natural unit are considered one of the major potential sources of new insights into the unfolding process, and more sophisticated statistical procedures and empirical studies will need to be conducted to fully exploit this potential. Two specific directions of research are suggested. In conjunction with the study of estimation procedures, the construction of latent classes in which the unit might be the same for all items within each class but different in different classes might be considered. If the locations of statements were the same in the classes, then the difference in the classes would be a matter of scale, as reflected in the different values of the unit. Alternatively, the unit may be constrained to be equal across classes but the statement locations may be free to vary. Then differences among classes would be interpreted to be qualitative and characterized by the different ordering of statements.

The unit parameter could be characterized as a person parameter rather than a statement parameter, or more completely, as a combined person and statement parameter that then could be resolved into a person and statement component such as $\theta_{ni} = \xi_n - \sigma_i$. This parameterization is beyond the scope of this paper.

The ramifications of the unit are especially interesting when juxtaposed with the PARELLA model proposed by Hoijtink (1990) and mentioned earlier as a generalization of the SSLM. The PARELLA model overcomes the limitation of the SSLM that the maximum probability of an endorsed response is only .5. In the generalization, as the person's location moves closer to the location of the statement, the probability of a response comes closer to 1.0 for increasing values of the "unit" parameter.

The 2PHCM has the same feature that the probability of an Agree response is not limited; furthermore, the probability of an endorsed response clearly becomes greater as the unit becomes longer for a given location of a person relative to a statement. Thus, it is clear how the parameters operate on the continuum to affect the probability. In contrast, it is not clear how to interpret the unit parameter directly on a continuum in the PARELLA model. It does not appear as clear as in the 2PHCM in which the unit is the distance of a threshold (in each direction) from the location of the statement for the probability of disagreeing is equal to the probability of agreeing. In addition, in

the PARELLA model, the unit parameter does not qualify each statement but only a set of statements.

Appendix: Derivation of Second Derivatives of the Likelihood Function

From Equation 23, which gives the expression for the first derivative,

$$\frac{\partial \log L}{\partial \delta_i} = \sum_n (x_{ni} - \pi_{ni}) \tanh(\beta_n - \delta_i) , \tag{54}$$

the second derivative is given by

$$\frac{\partial^2 \log L}{\partial^2 \delta_i} = \frac{\partial}{\partial \delta_i} \sum_n (x_{ni} - \pi_{ni}) \tanh(\beta_n - \delta_i)$$

$$= \sum_n x_{ni} \frac{\partial \tanh(\beta_n - \delta_i)}{\partial \delta_i} - \sum_n \pi_{ni} \frac{\partial \tanh(\beta_n - \delta_i)}{\partial \delta_i} - \sum_n \tanh(\beta_n - \delta_i) \frac{\partial \pi_{ni}}{\partial \delta_i}$$

$$= \sum_n (x_{ni} - \pi_{ni}) \frac{\partial \tanh(\beta_n - \delta_i)}{\partial \delta_i} - \sum_n \tanh(\beta_n - \delta_i) \frac{\partial \pi_{ni}}{\partial \delta_i} . \tag{55}$$

Because

$$E[x_{ni}] = \pi_{ni} E\left[\sum_n (x_{ni} - \pi_{ni}) \frac{\partial \tanh(\beta_n - \delta_i)}{\partial \delta_i} \right] = 0 , \tag{56}$$

$$E\left[\frac{\partial^2 \log L}{\partial^2 \delta_i} \right] = -\sum_n \tanh(\beta_n - \delta_i) \frac{\partial \pi_{ni}}{\partial \delta_i}$$

$$= -\sum_n \tanh(\beta_n - \delta_i) \pi_{ni}(1 - \pi_{ni}) \tanh(\beta_n - \delta_i)$$

$$= -\sum_n \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) . \tag{57}$$

Analogously,

$$E\left[\frac{\partial^2 \log L}{\partial \beta_n^2} \right] = -\sum_i \pi_{ni}(1 - \pi_{ni}) \tanh^2(\beta_n - \delta_i) . \tag{58}$$

References

Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology, 26,* 31–44.

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,* 561–573.

Andrich, D. (1979). A model for contingency tables having an ordered response classification. *Biometrics, 35,* 403–415.

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

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon Tuma (Ed.), *Sociological methodology* (pp. 33–80). San Francisco: Jossey-Bass.

Andrich, D. (1988). The application of an unfolding model of the PIRT type for the measurement of attitude. *Applied Psychological Measurement, 12,* 33–51.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement, 13,* 193–216.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgement and choice.* San Francisco: Holden Day.

Böckenholt, U., & Böckenholt, I. (1990). Modeling individual differences in unfolding preference data: A restricted latent class approach. *Applied Psychological Measurement, 14,* 257–269.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs 1: The method of paired comparisons. *Biometrika, 39,* 324–345.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences. Volume 1: Theory* (pp. 105–155). New York: Seminar Press.

Carroll, J. D., & De Soete, G. (1991). Toward a new paradigm for the study of multi-attribute choice behavior. *American Psychologist, 46,* 342–351.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review, 57,* 145–158.

Coombs, C. H., & Avrunin, C. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review, 84,* 216–230.

Coombs, C. H., & Smith, J. E. K. (1973). On the detection of structure in attitudes and developmental process. *Psychological Review, 80,* 337–351.

Cooper, L. G., & Nakanishi, M. (1983). Two logit models for external preferences. *Psychometrika, 48,* 607–620.

David, H. A. (1963). *The method of paired comparisons.* New York: Hafner.

Davidson, R. R., & Farquhar, P. H. (1976). A bibliography on the method of paired comparisons. *Biometrics, 32,* 241–252.

Davison, M. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika, 42,* 523–548.

DeSarbo, W. A., & Hoffman, D. L. (1986). Simple and weighted threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement, 10,* 247–264.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). New York: Wiley.

Guttman, L. (1954). The principal components of scalable attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 216–257). Glencoe IL: The Free Press.

Haberman, S. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics, 5,* 815–841.

Heiser, W. J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 3–31). Amsterdam: North-Holland.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika, 55,* 641–656.

Jansen, P. G. W. (1984). Relationships between the Thurstone, Coombs, and Rasch approaches to item scaling. *Applied Psychological Measurement, 8,* 373–383.

Jansen, P. G. W., Van den Wollenberg, A. L., & Wierda, F. W. (1988). Correcting unconditional parameter estimates in the Rasch model for inconsistency. *Applied Psychological Measurement, 12,* 297–306.

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

Poole, K. T. (1984). Least squares metric, unidimensional unfolding. *Psychometrika, 49,* 311–323.

Post, W. J., & Snijders, T. A. B. (1992). *Nonparametric unfolding models: A latent structure approach.* Leiden, The Netherlands: DSWO Press.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press.]

Schönemann, P. H. (1970). On metric multidimensional unfolding. *Psychometrika, 35,* 349–366.

Schönemann, P. H., & Wang, M. M. (1972). An individual difference model for the multidimensional analysis of preference data. *Psychometrika, 37,* 275–309.

Sixtl, F. (1973). Probabilistic unfolding. *Psychometrika, 38,* 235–248.

Tanzer, N. (1983). *Meassung von Einstellungen mittels probabilisticher Mebmodelle* [The measurement of attitudes by means of probabilistic measurement models]. Unpublished doctoral dissertation, University of Vienna.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology, 17,* 446–457.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34,* 278–286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33,* 529–554.

Van Schuur, W. H. (1984). *Structure in political beliefs, a new model for stochastic unfolding with application to European party activists.* Amsterdam: CT Press.

Van Schuur, W. H. (1988). Stochastic unfolding. In W. E. Saris & I. N. Gallhofer (Eds.), *Sociometric research. Volume 1: Data collection and scaling* (pp. 137–158). London: Macmillan.

Van Schuur, W. H. (1989). Unfolding German political parties: A description and application of multiple unidimensional unfolding. In G. De Soete, H. Ferger, & K. C. Klauer (Eds.), *New developments in psychological choice modelling* (pp. 259–277). Amsterdam: North Holland.

Verhelst, N. D., & Verstralen, H. H. F. M. (1991). *A stochastic model for inherently missing data.* Unpublished manuscript.

Wohlwill, J. F. (1963). The measurement of scalability for non-cumulative items. *Educational and Psychological Measurement, 23,* 543–555.

Wright, B. D. (1988). The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, van den Wollenberg, and Wierda. *Applied Psychological Measurement, 12,* 315–324.

Wright, B. D., & Douglas, G. A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement, 37,* 47–60.

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

Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika, 39,* 327–350.

Acknowledgments

Author's Address

Send requests for reprints or further information to David Andrich, School of Education, Murdoch University, Murdoch, Western Australia, Australia, 6150.