THE RATING SCALE MODEL FOR OBJECTIVE MEASUREMENT

Benjamin D. Wright and Graham A. Douglas

Memorandum No. 35
MESA Psychometric Laboratory
Department of Education
University of Chicago
February 1986

ABSTRACT

Necessary and sufficient relations between measurement objectivity and psychometric models for data in more than two categories are reviewed and extended. Rasch (1960, 1961) and others prove the sufficiency of Rasch models for objectivity. Douglas and Wright (1986) derive the model necessary for objectivity for data in two categories. Rasch (1968) outlines a proof of the model necessary for observations in any number of categories but does not deal with the unidimensional rating scale model. This paper completes Rasch's proof, interprets the structural characteristics of his model and shows that it is the only rating scale model which produces the measurement objectivity necessary for scientific comparisons.

Key words: item response theory, latent trait analysis, measurement, Rasch model, rating scale analysis

Introduction

Rasch (1960, 1961, 1967, 1977) identifies the necessary structure of unidimensional models for two observational categories. Douglas and Wright (1986) review and extend the proofs of this structure and address the related issue of parameter dimensionality. Our aim here is to derive the unidimensional model necessary for objectivity when observations are recorded in more than two categories.

Rasch's (1968) analysis of the necessary model when data are observed in more than two categories is unpublished and incomplete. With the exception of Andersen's (1972, 1973) work on parameter estimation, this multidimensional model has been neglected. But a considerable amount has been published on unidimensional Rasch models for rating scale data (Andrich 1978a, 1978b, 1978c; Masters 1982; Wright and Masters 1982; Masters and Wright 1984). Andrich (1978a) derives a unidimensional Rasch model for rating scale data from Thurstone's concept of thresholds. We will prove the necessity of this model from the requirement of objectivity.

The first section of this paper reviews the concept of objectivity (Douglas and Wright 1986). The second section proves the necessity of the Rasch rating scale model, i.e., one dimension but more than two categories of observation. The third section derives the necessary scoring function. The fourth section develops an interpretation of the category characteristics (Andrich 1978a).

Objectivity

Discussions of the essential part objectivity plays in scientific comparisons can be found in Rasch (1960, 1977) and Douglas and Wright (1986). A scientific comparison of objects with respect to a specific variable is a statement about the objects obtained from observations of their interactions with agents suitable to elicit manifestations of the variable in question. In order for a comparison to be objective, i.e., more than locally descriptive, it is necessary that the statement comparing the objects be independent of which agents have been employed to produce the observed interactions.

Interactions are realized through observational categories designed to define increasing levels of the intended variable. What we observe is the presence or absence of a particular event. This makes the observations qualitative. A model is needed to transform these qualitative observations into quantitative measures. When the model is probabilistic and unidimensional, the chances of an observation falling in a particular category must be determined wholly and uniquely by a pair of unidimensional parameters: one a property of the object and the other a commensurable property of the agent.

Since the probability of the observation is to be uniquely determined by the property common to object O and agent A, a general expression for a comparison of two objects O_1 and O_2 with parameters β_1 and β_2, based on a suitable agent A with parameter δ is

$$g\left[P(\beta_1,\delta),P(\beta_2,\delta)\right] .$$

Douglas and Wright (1986) show that when there are two categories the only probability function P (i.e., model) that can make the comparative statement g independent of δ is

$$P(\beta,\delta) = \exp\left[f_1(\beta)+f_2(\delta)+C\right] / \left[1+\exp\left(f_1(\beta)+f_2(\delta)+C\right)\right] . \qquad (1)$$

The crucial feature of this model is that the function of the object parameter $f_1(\beta)$ and the function of the agent parameter $f_2(\delta)$ enter the exponent additively. The familiar Rasch model for measurement based on two categories arises by replacing $f_1(\beta)$ by β, $f_2(\delta)$ by $-\delta$, and setting $C = 0$. An indispensable aspect of the derivation of this result is the statistical conditioning which makes comparisons objective by removing unwanted parameters.

Models for More than Two Categories

A set of $m + 1$ categories sustains m pairs of object and agent parameters, each pair representing a different dimension (Douglas and Wright 1986). A unidimensional model for rating scale data must focus on one of these pairs. We will describe the participating elements as "objects" and "agents" to emphasize that the rating scale model is applicable to any situation in which the observation of the interaction between an object and an agent falls into one of a set of categories. The unidimensional framework means that there can be one (and only one) parameter per object , i.e., one β , and one (and only one) commensurable parameter per agent, i.e., one δ . It also means that the categories can have one and only one order from "least " to "most." Indeed, establishing that order by assigning hierarchical labels to the categories is the basic step in operationalizing the intended variable.

Our proof will show that a useful model exists with these characteristics and that it is the only model for such a framework which has objectivity. Psychometric applications follow from suitable identification of objects, agents and categories. We will also show that à priori "scoring" of categories is not only superfluous but out of order, since one and only one way of counting observations results from the model's structure.

Rating Scale Model

When observations are made in $m + 1$ ordered categories, we can write $U:(X^0,X^1,..,X^m)$ to represent the set of available categories. Within this complete set of $m + 1$ possibilities we can represent partitions which isolate pairs of categories by $V:(X^0,X^j)$.

A general model for a set of categories U is

$$P(X \varepsilon X^j) = P_j(\beta,\delta) \qquad j=0,1,..,m, \tag{2}$$

where (i) the notation $X \varepsilon X^j$ means that observation X falls in category X^j ,

(ii) the index j on $P_j(\beta,\delta)$ reminds us that the probability will depend on the category X^j in which the observation fell, and

(iii) objects and agents are each parameterized by a single scalar parameter to conform to the unidimensional framework.

A unidimensional framework requires that β and δ combine in $P_j(\beta,\delta)$ in one and only one way because more than one way of combining would produce more than one dimension. This means that β and δ must combine in a function $\mu(\beta,\delta)$ which does not change from category to category, i.e., which is not a function of j . This makes the model

$$P(X \varepsilon X^{j}) = P_{j}\left[\mu(\beta,\delta)\right] \qquad j=0,1,..,m. \tag{3}$$

Our task is to derive the structure of $P_{j}\left[\mu(\beta,\delta)\right]$ necessary for objectivity. When the framework is partitioned so that attention is focused on a subset V of two categories within U , a subset in which the observation X is recorded only if it belongs to one of the two categories within V , then the probability that the observation is in category X^{j} , given that it came from V , is

$$P(X \varepsilon X^{j}|V) = \frac{P(X \varepsilon X^{j}|U)}{P(V|U)} \quad . \tag{4}$$

A similar expression may be written for the probability that the observation is in category X^{0} within V . When we take the ratio of these two probabilities, their denominators cancel leaving

$$\frac{P(X \varepsilon X^{j}|V)}{P(X \varepsilon X^{o}|V)} = \frac{P(X \varepsilon X^{j}|U)}{P(X \varepsilon X^{o}|U)} \quad . \tag{5}$$

This is remarkable because it shows that the ratio of probabilities of pairs of observations is the same regardless of the partition of the framework. Thus, if we know the structure of the probabilities with respect to any pair in any V, we also know the structure of all pairs in U.

Since the structure necessary for objectivity within V is known to be the additive exponential of equation (1) (Douglas and Wright 1986), the ratio of probabilities of the pair of categories X^j and X^0 must also be an additive exponential of some scalar functions of the object parameter, say $\zeta_j(\beta)$ and of the agent parameter, say $\psi_j(\delta)$. The ratio of probabilities in (5) must therefore be

$$\frac{P(X \in X^j | U)}{P(X \in X^o | U)} = \exp\left[\zeta_j(\beta)+\psi_j(\delta)\right]. \tag{6}$$

Since neither unidimensional parameter β nor δ may be indexed to categories, the functions ζ_j and ψ_j are subscripted by j to identify the category they refer to.

The remainder of the derivation of the rating scale model necessary for objectivity is accomplished by manipulations of these functions and differentiation. The crux of the derivation is that objectivity requires the unidimensional function which combines object and agent parameters $\mu(\beta,\delta)$ to be the same from one category to another.

We start by replacing $P(X \epsilon X^j)$ with $P_j\left[\mu(\beta,\delta)\right]$ in the ratio (6) and removing the exponential by taking logarithms,

$$\log\left[\frac{P_j\left[\mu(\beta,\delta)\right]}{P_o\left[\mu(\beta,\delta)\right]}\right] = \zeta_j(\beta)+\psi_j(\delta) \ . \tag{7}$$

This exposes ζ_j and ψ_j for analysis. Letting $g_j\left[\mu(\beta,\delta)\right]$ represent this ratio, we have

$$g_j\left[\mu(\beta,\delta)\right] = \zeta_j(\beta)+\psi_j(\delta) \ . \tag{8}$$

Our aim is to obtain an additive structure for object and agent parameters on both sides of this equation. This is done by reparameterizations based on the functional relationships among these parameters.

The first step is to reparameterize category one by setting $\zeta_1(\beta)$ equal to θ and $\psi_1(\delta)$ equal to η , so that

$$g_1\left[\mu(\beta,\delta)\right] = \zeta_1(\beta)+\psi_1(\delta) = \theta+\eta \ . \tag{9}$$

But now the parameters on the left are unlike those on the right. This disparity is remedied by inverting the function in (9) so that

$$\mu(\beta,\delta) = g_1^{-1}(\theta+\eta) = k_1(\theta+\eta) \; . \tag{10}$$

When this expression for μ is introduced into (8) for category X^2 , we have

$$g_2\left[k_1(\theta+\eta)\right] = \zeta_2(\beta)+\psi_2(\delta) \; . \tag{11}$$

Once again we have a disparity in parameters on left and right. But both sets are additive. The disparity is removed as before. From (9) we write

$$\beta = \zeta_1^{-1}(\theta) = s_1(\theta) \quad \text{and} \quad \delta = \psi_1^{-1}(\eta) = t_1(\eta) \tag{12}$$

so that $\zeta_2(\beta)+\psi_2(\delta)$ may be written

$$\zeta_2\left[s_1(\theta)\right] + \psi_2\left[t_1(\eta)\right] \; , \tag{13}$$

and the functions-of-a-function may be replaced by single functions, say

$$\zeta_2\left[s_1(\theta)\right] = h_2(\theta) \quad \text{and} \quad \psi_2\left[t_1(\eta)\right] = k_2(\eta) \tag{14}$$

so that (11) becomes

$$g_2\left[k_1(\theta+\eta)\right] = h_2(\theta)+k_2(\eta) \quad . \tag{15}$$

Now the parameters θ and η are the same on both sides. The function g_2 for category X^2 is a function of k_1 . But k_1 is a function of $\theta+\eta$ which is not associated with the second category. Finally g_2 is equal to an additive combination of functions h_2 and k_2 of the same θ and η , but indexed by $j = 2$.

Partial derivatives of the function $(\theta+\eta)$ with respect to either θ or η are equal to 1 . To take advantage of this we differentiate the function of a function of a function in (15) with respect to θ and then separately with respect to η . Thus

$$\frac{\partial g_2}{\partial k_1}\,\frac{\partial k_1}{\partial(\theta+\eta)}\,\frac{\partial(\theta+\eta)}{\partial\theta} = \frac{\partial h_2(\theta)}{\partial\theta}$$

and

$$\frac{\partial g_2}{\partial k_1}\,\frac{\partial k_1}{\partial(\theta+\eta)}\,\frac{\partial(\theta+\eta)}{\partial\eta} = \frac{\partial k_2(\eta)}{\partial\eta} \quad . \tag{16}$$

But the left sides of these expressions are equal because the first two differentiations in each expression are identical and the derivatives

$$\frac{\partial(\theta+\eta)}{\partial\theta} \quad \text{and} \quad \frac{\partial(\theta+\eta)}{\partial\eta} \quad \text{equal } 1 \ .$$

This shows that the derivative of h_2 with respect to θ is identical to the derivative of k_2 with respect to η for all values of θ and η .

The only way these derivatives can be identical is for each to be equal to the same constant indexed by $j = 2$. When we set each derivative equal to this constant, say ϕ_2 , we have two differential equations,

$$\frac{\partial h_2(\theta)}{\partial\theta} = \phi_2 \qquad \text{and} \qquad \frac{\partial k_2(\eta)}{\partial\eta} = \phi_2 \ . \tag{17}$$

The solutions to these equations are

$$h_2(\theta) = \omega_2+\phi_2\theta$$

and

$$k_2(\eta) = \chi_2+ \phi_2\eta$$

in which the constants ω_2 and χ_2 are not distinguishable and can be combined into one constant κ_2. This gives (15) the form

$$g_2\left[k_1(\theta+\eta)\right] = \kappa_2+\phi_2(\theta+\eta) . \tag{18}$$

When we repeat these steps for any category j we find that the function $k_1(\theta+\eta)$ reappears each time and that for the general category x^j we have

$$g_j\left[k_1(\theta+\eta)\right] = \kappa_j+\phi_j(\theta+\eta) \qquad j=1,2,..,m. \tag{19}$$

The last step is to restore the original parameters β and δ by means of (9) so that

$$\frac{P_j\left[\mu(\beta,\delta)\right]}{P_o\left[\mu(\beta,\delta)\right]} = \exp\left[\kappa_j+\phi_j\left(\zeta(\beta)+\psi(\delta)\right)\right] \quad . \tag{20}$$

The functions ζ and ψ of object and agent parameters enter additively into the rating scale model. These object and agent parameters may be transformed in any way we wish as long as the transformations enter additively. For psychometric applications where the agent is a rating scale item, the item's characteristic might be labeled "affective value"--the item's tendency to attract responses into categories towards the "more" end of the ordered response set. Semantic problems are unlikely to arise, however, if we continue to use the term "item difficulty" and replace $\zeta(\beta)$ by β and $\psi(\delta)$ by $-\delta$, producing

$$\frac{P(X \varepsilon X^j)}{P(X \varepsilon X^o)} = \exp\left[\kappa_j+\phi_j(\beta-\delta)\right] \tag{21}$$

as the form of (6) necessary for objectivity.

By shifting the denominator in (21) to the right, summing over j and setting the sum of all category probabilities to 1, we produce an expression for the probability of the observation being in the first category,

$$P(X \varepsilon X^0) = \frac{\exp\left[\kappa_0+\phi_0(\beta-\delta)\right]}{\sum_{k=0}^{m}\exp\left[\kappa_k+\phi_k(\beta-\delta)\right]} . \tag{22}$$

In general we have

$$P(X \varepsilon X^j) = \frac{\exp\left[\kappa_j+\phi_j(\beta-\delta)\right]}{\sum_{k=0}^{m}\left[\exp\ \kappa_k+\phi_k(\beta-\delta)\right]} , \quad j=0,1,\ldots,m , \tag{23}$$

Since this model must describe the framework for any m, it must do so for the case of $m = 1$, the dichotomous model. This shows that the values of κ_0, κ_m and ϕ_0 can be zero. But it does not tell us their values when m is greater than 1. Thus when $m = 2$, we can set $\kappa_0 = \kappa_2 = \phi_0 = 0$ but we need values for κ_1, ϕ_1 and ϕ_2.

Scoring Functions

A rating scale model was derived from the requirement of objectivity. In the process we had to introduce two characteristics κ and ϕ. Since κ and ϕ do not describe objects or agents, how they relate to objectivity is not obvious. To clarify this we will investigate what happens to κ and ϕ when we estimate δ.

We commence by replacing the observation X by a vector (a_{nij}) which uses the label 1 to indicate when object n's interaction with agent i is in category j, and 0 otherwise. The interaction of one object with one agent becomes a string of indicators, the observation vector (a_{nij}), in which all entries except one are zero. The exception has the indicator 1 in position j. The model is

$$P\left[(a_{nij})\right] = \exp\left[\sum_{j=o}^{m} a_{nij}\left[\kappa_j+\phi_j(\beta_n-\delta_i)\right]\right] / \gamma_{ni} \qquad (24)$$

where (i) (a_{nij}) is the vector of zeroes and a one, and
(ii) γ_{ni} is the sum of the $m + 1$ numerators.

Estimation of the δ's commences by forming the probability of a set of observations when any object is exposed to L agents,

$$P\left[((a_{nij}))\right] = \prod_{i=1}^{L}\left[\frac{\exp \sum_{j=0}^{m} a_{nij}\left[\kappa_j+\phi_j(\beta_n-\delta_i)\right]}{\gamma_{ni}}\right]$$

$$= \exp\left[\sum_{j=0}^{m} T_{nj}(\kappa_j+\phi_j\beta_n)-\sum_{i=1}^{L}\left(\sum_{j=0}^{m} a_{nij}\phi_j\right)\delta_i\right] \Big/ \prod_{i=1}^{L}\gamma_{ni} \qquad (25)$$

where (i) T_{nj} is the number of times the interaction between object n and the L agents fell in category j ,

(ii) κ_j appears as a linear term separate from β_n and δ_i ,

(iii) ϕ_j multiplies T_{nj} in the expression involving β_n , and

(iv) the coefficient of δ_i is a product of the indicator a_{nij} and ϕ_j .

The objective estimation of the δ's requires the identification of a statistic which conditions out parameters such as β and unknown characteristics such as κ and ϕ . The way to accomplish this is to find such a statistic, derive its probability distribution, and form the conditional distribution of the data given the statistic. This is done by dividing the unconditional distribution of the data (25) by the distribution of the statistic.

The way data bear on κ and ϕ in (25) through their multiplication by the category counts T_{nj} suggests that the vector of category counts (T_{nj}) is the statistic necessary to condition out κ and β. To get the distribution of this vector (T_{nj}) we sum the probability (25) over all ways in which observations $((a_{nij}))$ could add up to this set of (T_{nj}). This summation is

$$P\left[(T_{nj})\right] = \sum_{((a_{nij}))}^{(T_{nj})} P\left[((a_{nij}))\right]$$

$$= \sum_{((a_{nij}))}^{(T_{nj})} \exp\left[\sum_{j=0}^{m} T_{nj}(\kappa_j + \phi_j \beta_n)\right] \exp\left[-\sum_{i=1}^{L} (\sum_{j=0}^{m} a_{nij}\phi_j)\delta_i\right] / \prod_{i=1}^{L} \gamma_{ni} \, . \tag{26}$$

The numerator of (26) is expressed as a product of two parts. In the second part, the data a_{nij} remain uncombined. But in the first part the data are combined into the T_{nj}. Because (T_{nj}) does not change within summations over data which satisfy (T_{nj}), the first part is a constant and can be moved outside that summation.

The final step in deriving the conditional distribution is to form the ratio of the distributions, (25) and (26). In this ratio the expressions

$$\exp\left[\sum_{j=0}^{m} T_{nj}(\kappa_j+\phi_j\beta_n)\right] \text{ and } \prod_{i=1}^{L} \gamma_{ni} \quad ,$$

appearing in numerator and denominator cancel, so that the remaining conditional distribution is a function of a_{nij}, ϕ_j and δ_i only.

The ϕ's are the final obstacle to objectivity. No further conditioning can remove them. The way to discover their structure is to analyze the manner in which ϕ_j combines with T_{nj} in (25) and (26) to form the object score,

$$R_n = \sum_{j=0}^{m} \phi_j T_{nj} \quad . \tag{27}$$

We can see what this scoring tells us about the ϕ's if we review what the ordering of categories implies about the relationship between data and the measures derived from them.

The initiating definition of any variable implies a continuum from "less" to "more." Variables are operationalized to obtain evidence of this continuum by labeling their categories in such a way that the labels establish the intended direction of "more." This produces monotonicity between the intended direction of the variable, the hierarchy of category labels and measures on the variable. Just as "right" answers are defined as smarter than "wrong" answers, observations in higher categories are defined to indicate "more" of the variable than those in lower categories.

Since R_n is the score from which the measure will be inferred, we can learn about the ϕ's by considering scores differing by a one-category improvement, say from category p up to category $p+1$ as in

$$\begin{aligned} R_n &= \ldots+\phi_p T_{np}+\phi_{p+1} T_{n,p+1}+\ldots \\ R_n' &= \ldots+\phi_p \left[T_{np}-1\right]+\phi_{p+1}\left[T_{n,p+1}+1\right]+\ldots \quad . \end{aligned} \tag{28}$$

Because this is an improvement, it follows that

$$R_n' > R_n \qquad \text{or} \qquad (R_n' - R_n) > 0 \ .$$

But by taking the difference between R_n' and R_n in (28) we have

$$R_n'-R_n = \phi_{p+1}-\phi_p > 0 \quad .$$

This shows the ϕ's must have the same order as the category labels, i.e.,

$$\phi_0 < \phi_1 < \cdots < \phi_m \quad . \tag{29}$$

Now consider a third R_n'' formed by any other one-category improvement over R_n, say from category q up to category $q + 1$. Differences $R_n' - R_n$ and $R_n'' - R_n$ are one-category improvements from the same R_n. Unidimensionality requires that one-category improvements increase scores in such a way that no possible scores are skipped. Objectivity requires that the effect on a measure of any one-category improvement from R_n (whether to R_n' or to R_n'') be independent of which agents or categories are involved. This means that $R_n' - R_n$ and $R_n'' - R_n$ must be equal. Thus

$$\phi_{p+1} - \phi_p = \phi_{q+1} - \phi_q \quad , \text{ all } p,q \; . \tag{30}$$

Since this difference is independent of which categories are involved, it must be a constant,

$$\phi_{p+1} - \phi_p = C \; . \tag{31}$$

If we set ϕ_0 equal to zero (as we do when $m = 1$), the succeeding ϕ's may be written

$$\phi_1 = C \ , \ \phi_2 = 2C \ , \ .. \ , \ \phi_m = mC \ , \tag{32}$$

This shows that C is an arbitrary scaling unit which may as well be 1 .

Thus the particular scoring function necessary for objectivity, in its simplest form, is

$$\phi_j = j \quad , \quad j=0,1,..,m \quad . \tag{33}$$

The necessary model becomes

$$P(X \ \varepsilon \ \mathbf{X}^j) = \exp\left[\kappa_j + j(\beta_n - \delta_i)\right] / \gamma_{ni} \ . \tag{34}$$

The score

$$R_n = \sum_{j=0}^{m} jT_{nj}$$

produced by the ϕ's is just the count of the number of categories surpassed by (i.e., below) each observation. This counting stems from the ordering of the categories determined by the definition of the variable.

Category Structure and Multidimensionality

The remaining unknowns in the rating scale model for objective comparisons of objects and agents in one dimension are the κ's . The absence of object or agent subscripts suggests that the κ's are structural parameters (Andrich 1978a, 1985). However, the probability of the interaction of one object with one agent

$$P\left[(a_{nij})\right] = \exp\left[\sum_{j=o}^{m} a_{nij}\left[\kappa_j + j(\beta_n - \delta_i)\right]\right] / \gamma_{ni} ,$$

shows that the κ's do not multiply objects or agents like the ϕ's . Instead they are additive like object and agent parameters. Since the object and agent parameters on the one dimension of the rating scale model are exhausted by β and δ , the κ's must either be category descriptors or represent the object and agent parameters of a model with more than one dimension.

In frameworks of $m + 1$ categories there are m possible dimensions and the dimensionality of objects and agents must be identical (Douglas and Wright 1986). This means that when we want two parameters to characterize any object, we must have two for each agent which interacts with that object and, also, observations in at least three categories.

When we apply these conditions to the rating scale model with m categories but only one dimension, we see that $m - 1$

additional dimensions corresponding to $m - 1$ different arrangements of the $m + 1$ category labels could also be modeled. Rasch (1968) proves that the multidimensional model necessary for objectivity can be written

$$P\left[(a_{nij})\right] = \exp\left[\sum_{j=0}^{m} a_{nij}(\beta_{nj}-\delta_{ij})\right] / \gamma_{ni}, \tag{35}$$

where (i) a_{nij} is as before,

(ii) β_{nj} is the parameter for object n on dimension j,

(iii) δ_{ij} is the parameter for agent i on dimension j, and

(iv) a convenient choice of restrictions to bring $m + 1$ categories into m dimensions is is $\beta_{no} = \delta_{io} = 0$.

Apart from additional object and agent parameters, the difference between the rating scale model for one dimension and this model for m dimensions is the absence of ϕ's and κ's. We can have a scoring function in (35) by introducing ϕ_{kj} as a multiplier for

$$\phi_{kj}(\beta_{nj}-\delta_{ij}) \quad \text{with} \quad \phi_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} .$$

The result resembles that of Andrich (1985) but differs in that his model has less than m dimensions so that his $((\phi_{kj}))$ cannot be an identity matrix.

This brings us to the κ's and the part they play with respect to extra dimensions. One way to see what the κ's represent is to separate a "first" dimension from the m dimensions of the m-dimensional model. This leads to

$$\beta_{nj}-\delta_{ij} = \sum_{k=1}^{m}\phi_{kj}(\beta_{nk}-\delta_{ik})$$

$$= \phi_{1j}(\beta_{n1}-\delta_{i1})+ \sum_{k=2}^{m}\phi_{kj}(\beta_{nk}-\delta_{ik})$$

which, when we score the first dimension by $\phi_j = j$, becomes

$$\beta_{nj} -\delta_{ij} = j(\beta_{n1}-\delta_{i1})+ \sum_{k=2}^{m}\phi_{kj}(\beta_{nk}-\delta_{ik}) .$$

Since the m and unidimensional models account for data in the same $m + 1$ categories, the counterpart for the unidimensional model

$$\kappa_j +j(\beta_n-\delta_i)$$

shows that the κ's replace $m - 1$ unmodeled parameters via

$$\kappa_j = \sum_{k=2}^{m} \phi_{kj}(\beta_{nk}-\delta_{ik}) \qquad j=0,1,..,m \; . \tag{36}$$

When we model a second dimension, the same principle of reduction applies leaving

$$\kappa_j = \sum_{k=3}^{m} \phi_{kj}(\beta_{nk}-\delta_{ik}) \qquad j=0,1,..,m \; .$$

Thus, for models of dimension less than m, there must always be κ's to summarize the remaining dimensions potential in the observations but not modeled. But when step-wise extraction of dimensions reaches m, the explicit object and agent parameters exhaust the parameter structure leaving no dimensions for the κ's to represent.

Another way to see this is to express the connection between κ_j and the m-dimensional model as

$$\kappa_j = (\beta_{nj}-j\beta_n)-(\delta_{ij}-j\delta_i) \tag{37}$$

in which (i) β_{nj} and δ_{ij} represent the dimension implied by a comparison of the jth category with all others, and

(ii) β_n and δ_i are the parameters of the rating scale model in which categories are scored $0,1,..,m$.

This shows how each κ_j summarizes the effect of the unmodeled dimensions on the use of the jth category. Since the κ's confound the contributions of unmodeled object and agent parameters, they cannot be estimated objectively. We can expect κ_j to approximate invariance only in situations in which $\beta_{nj} - j\beta_n$ and $\delta_{ij} - j\delta_i$ approximate constants over the n and i involved. While different samples within a given framework may produce objective estimates of the modeled β and δ, the effect of unmodeled dimensions on the κ's will, in general, cause them to vary from sample to sample.

Constructing Multidimensionality

Since objects and agents may be estimated independently of κ's , it follows that any one dimension can be analyzed independently of any other dimensions which could arise from the category structure. This means that the m dimensions in $m + 1$ categories can be constructed one at a time.

We have shown that consecutive integers, related monotonically to the hierarchy of categories defining a dimension, provide the scoring necessary for objectivity in that dimension. The hierarchy of categories comes from the particular ordering of category labels which implements the intended dimension. But this must be so for any intended dimension. Therefore the construction of additional dimensions is equivalent to reordering categories.

This enables the construction of multidimensional frameworks one dimension at a time. For each dimension we reorder the category labels to define the intended variable and represent each category ordering with successive integers of the form

$$\phi_j = k \quad , \quad k = 0,1,..,n \quad , n \leqslant m \quad .$$

For example, the category labels "Strongly Disagree," "Disagree," "Agree," and "Strongly Agree" are frequently used in frameworks designed to elicit one dimension of attitude. The Likert scoring of these four labels with the integers 1, 2, 3 and 4 (i.e., 0, 1, 2 and 3) produces person and item estimates along a variable, "Amount of Attitude." One parameter, β_{n1} for person attitude, and one parameter, δ_{i1} for item affectivity, suffice to describe the location of people and items on this dimension.

A second scoring function

$$\phi_0 = 1 \text{ , } \phi_1 = 0 \text{ , } \phi_2 = 0 \text{ and } \phi_3 = 1$$

reorders the category labels so that Strongly Disagree and Strongly Agree are combined to represent "More Strongly" and Agree and Disagree are combined to represent "Less Strongly." This reordering constructs a second dimension with a different interpretation, e.g., "Intensity of Response" (Guttman 1950; Andrich 1985), and with additional person parameter β_{n2} and additional item parameter δ_{i2} to locate people and items on this second dimension.

Objectivity of measurement requires that the interpretation of any one dimension be free of all aspects of other potential dimensions. This means that equivalent estimates $(\hat{\beta}_{n1}, \hat{\beta}_{n2})$ for any person and $(\hat{\delta}_{i1}, \hat{\delta}_{i2})$ for any item must result whether we analyze Intensity first and then Amount, or vice versa.

Estimation

When we used the ratio of (25) to (26) to derive the conditional distribution necessary for the objective estimation of item parameters, we found that the m statistics T_{nj} removed the β's and κ's from the model. The m agent count statistics S_{ij} provide a similar conditional distribution for the objective estimation of object parameters. This means that the

κ's need play no part in parameter estimation. On the other hand, routine application of a conditional estimation algorithm is not always the most efficient estimation method (Wright and Douglas 1977). Unconditional procedures (Wright and Masters 1982) for parameter estimation in the rating scale model, however, necessitate the calculation of κ's along with the estimates of β and δ.

The unconditional likelihood of a set of data when $m = 1$ is

$$P\left[(((a_{nij})))\right] = \exp\left[\sum_{j=0}^{m} T_j \kappa_j + \sum_{n=1}^{N} R_n \beta_n - \sum_{i=1}^{L} S_i \delta_i\right] / \prod_{n=1}^{N} \prod_{i=1}^{L} \gamma_{ni} \quad . \tag{38}$$

This shows that the set (T_j) is sufficient for (κ_j). But

(i) $\sum_{j=0}^{m} T_j = NL$, i.e., the sum of the T's is equal to the total number of interactions between N objects and L agents, and

(ii) $\sum_{j=0}^{m} jT_j = \sum_{n=1}^{N} R_n = \sum_{i=1}^{L} S_i$, i.e., the sum of the category scores is equal to the sum of the object scores and to the sum of the agent scores.

Thus two restrictions are necessary to calculate a set of κ's .

Convenient restrictions which maintain consistency with the two category ($m = 1$) model are $\kappa_0 = \kappa_m = 0$.

Summary and Conclusions

This paper applies the principle of objectivity in scientific comparisons to unidimensional measurement models for data collected from a set of more than two categories. Objectivity in a model means that the magnitude of a measure estimated from that model is not affected in any important way by aspects of the framework other than the object itself. The model makes the measure independent of which agents are used to produce it, of what other objects may or may not have been measured and of any other elements of the framework necessary to effect the interaction between object and agent.

The requirement of objectivity as necessary for measurement leads to proofs identifying models with this property. Douglas and Wright (1986) review and extend these proofs when there are two categories. Rasch (1968) outlines a proof deriving the model necessary for objectivity for observations in $m + 1$ categories when m dimensions are modeled. In this paper we derive the necessary unidimensional model for observatons in $m + 1$ categories.

The necessary rating scale model has one and only one pair of object and agent parameters which describes objects and agents on a single common dimension. The log-odds of increasingly indicative events is governed entirely by a linear function of these object and agent parameters. This linear function is multiplied by a "scoring" coefficient the values of which must be known before objectivity can be obtained. A proof based on objectivity shows that the numerical values of this coefficient must be equivalent to the integers $0,1,..,m$. Andersen (1977) reaches the same conclusion with a proof based on the requirement of minimal sufficient statistics.

The identification of the scoring coefficient necessary for objectivity increases our understanding of the way in which qualitative observations can be converted into objective measures. Category scoring is always equivalent to counting steps taken upward through an ordered sequence of increasingly indicative events (Wright and Masters 1982).

The rating scale model necessary for objectivity also contains a set of category descriptors. Andrich (1978a, 1985) interprets these characteristics as category thresholds and discusses their relationship to Guttman's (1950) components of a scale. We derive an interpretation of these characteristics from Rasch's 1968 multidimensional model. When less than the maximum number of dimensions is modeled, the category characteristics summarize the unmodeled object and agent parameters. As more dimensions are modeled, the category characteristics represent less and less object and agent variation until, when m dimensions are modeled, no category characteristics are required.

Since the definition of a variable always depends on an ordering of category labels, the m dimensions observed in $m + 1$ categories can be accessed one dimension at a time by reordering categories. The scoring of each variable is implemented by applying the integers $0,1,..,n \leq m$ to the particular order of categories which defines that particular variable.

The significance of these Rasch models goes beyond psychometrics. They have been used to solve problems in sociology, anthropology, political science, archeology, ecology, criminology, civil engineering and biology. They are the only models which fulfill Thurstone's requirement for a unit-preserving process (1931, 257) which is sample-free (1928, 416) and, stochastically, Guttman's (1944) requirement that the response pattern be recoverable from the measure. Their mathematical structure produces the "additive conjoint" (Luce and Tukey 1964) and hence "fundamental" (Norman Campbell 1920) measurement which is the sine qua non of physical science. The objectivity which their structure makes possible is at the core of scientific inference and a basic premise in the theory of knowledge.

References

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. _Journal of the Royal Statistical Society_, _34_ 42-54.

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. (1978a). A rating formulation for ordered response categories. _Psychometrika_, _43_, 561-573.

Andrich, D. (1978b). Scaling attitude items constructed and scored in the Likert tradition. _Educational and Psychological Measurement_, _38_, 665-680.

Andrich, D. (1978c). Application of a psychometric rating model to ordered categories which are scored with successive integers. _Applied Psychological Measurement_, _2_, 581-594.

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.

Campbell, N. R. (1920 [1957]). _Physics: The elements_. London: Cambridge University Press. [Reprinted as _Foundations of Science_. New York: Dover, 1957.]

Douglas, G. A. and Wright, B. D. (1986). _The two category model for objective measurement_. (Research Memorandum No. 34). Chicago: University of Chicago, MESA Psychometric Laboratory.

Guttman, L. (1944). A basis for scaling qualitative data. _American Sociological Review_, _9_, 139-150.

Guttman, L. (1950). The principal components of a scale analysis. In Stouffer et al. (Eds.), _Measurement and prediction_ (pp. 312-361). New York: Wiley.

Luce, R. D. and Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. _Journal of Mathematical Psychology_, _1_, 1-27.

Masters, G. N. (1982). A Rasch model for partial credit scoring. _Psychometrika_, _47_, 149-174.

Masters, G. N. and Wright, B. D. (1984). The essential process in a family of measurement models. _Psychometrika_, _49_, 529-544.

Rasch, G. (1960 [1980]). _Probabilistic models for some intelligence and attainment tests_. [Reprint, with a Foreword and Afterword by Benjamin D. Wright. Chicago: University of Chicago Press, 1980.]

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. _Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability_ (pp. 321-333). Berkeley: University of California Press.

Rasch, G. (1967). An informal report on a theory of objectivity in comparisons. In L. J. Th. van der Kamp and C. A. J. Vlek (Eds.), Measurement Theory, Proceedings of the NUFFIC International Summer Session in Science in "Het Oude Hof," (The Hague, July 14-28, 1966, University of Leiden).

Rasch, G. (1968, September). A mathematical theory of objectivity and its consequences 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.

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

Thurstone, L. L. (1931). Measurement of social attitudes. Journal of Abnormal and Social Psychology, 26, 249-269.

Wright, B. D. and Douglas. (1977). Conditional versus unconditional procedures for sample-free item analysis. Educational and Psychological Measurement, 37, 573-586.

Wright, B. D. and Masters, G. N. (1982). Rating scale analysis. Chicago: MESA Press.