Relationships Between the Thurstone, Coombs, and Rasch Approaches to Item Scaling

1 author:

Paul Jansen
Vrije Universiteit Amsterdam

111 PUBLICATIONS **2,722** CITATIONS

Some of the authors of this publication are also working on these related projects:

From VU to Work View project

Ambition at work View project

Relationships Between the Thurstone, Coombs, and Rasch Approaches to Item Scaling

Paul G. W. Jansen
The Netherlands Postal and Telecommunications Services

Andrich (1978) derived a formal equivalency between Thurstone's Case V specialization of the law of comparative judgment for paired comparisons, with a logistic function substituted for the normal, and the Rasch model for direct responses. The equivalency was corroborated by a specific substantial-psychological interpretation of the Rasch binary item response probability. Studying the relationships between the Thurstone and Rasch models from another perspective than Andrich's, namely, from a data-theoretical point of view, it appears that the equivalency is based on an implicit assumption with respect to the subject population. This assumption (1) is rather restrictive, and therefore its empirical validity seems to be low, and (2) seems to contradict the substantial reasoning corroborating the Thurstone-Rasch equivalency. It is argued that the Thurstone model cannot be considered the sample-independent pair comparison counterpart of the Rasch model. An alternative pair comparison equivalent of the Rasch model is tentatively proposed. Finally, the theoretical and practical implications of Andrich's and of the present study are discussed.

Andrich (1978) demonstrated that item parameter estimates obtained from either a Thurstone Case V analysis of pair comparison data or a Rasch analysis of direct response data, may, under certain circumstances, be equivalent. In the former approach, subjects are required to choose from a pair (i,j) of items the item they prefer more with respect to the trait in question, and the basic data consist

of the $\binom{K}{2}$ observed choice proportions $p(i,j)$ representing the relative number of times that item i $(i = 1, \ldots, K)$ has been preferred to item j in the total sample of subjects. In the latter approach, subjects are asked to respond either positively (e.g., correct, agree, like) or negatively (e.g., incorrect, disagree, do not like) to an item, and the basic data consist of the $N \times K$ binary indicator variables a_{vi} $(v = 1, \ldots, N;$ subject index) representing a positive response $(a_{vi} = 1)$ or a negative response $(a_{vi} = 0)$ to the item with respect to the trait at issue.

In this paper, the equivalency is studied from a formal, or theoretical, perspective. Since the equivalency is valid given a specific condition (which is derived below), it is an equivalency *with respect to a certain aspect* of both models. In this way, model equivalency does not coincide with model identity, which would be trivial. However, this implies that there is something arbitrary in the choice of this "equivalency aspect": Choosing another aspect would result in another kind of equivalency, or even in apparent nonequivalency.

Therefore, Andrich's (1978) reasoning may be studied theoretically by contrasting its basic equivalency perspective to another equivalency point of view, which is done in this paper. The procedure is that, again, some sort of link between the Thurstone and Rasch models is searched for, but from another perspective than Andrich's. The new perspective is introduced under the heading "data-

theoretical'' in Section 2. It appears that, in terms of the theory of data, a possible bridge to close the gap between the Thurstone and Rasch models might be found in Coombs' (1964) unfolding model, and that this result sheds a new and restrictive light on Andrich's equivalency by making explicit a specific assumption underlying the latter (Section 3). To complete the discussion, in Section 4 it is investigated whether the perspective dominating Andrich's equivalency study may be justly denoted as ''sample independence'' (Andrich, 1978, p. 458). Finally, it is argued why and how theoretical model comparisons, as embodied by the study of Andrich and by this paper, are relevant to applied measurement (Section 5).

1. Andrich's Equivalency

The Thurstone (Case V) Model

In the Thurstone Case V model, the pairwise choice probability $\pi(i,j)$—the probability that a subject v, sampled at random from the reference population at issue, will choose item i from the pair (i,j)—is represented by $\Phi(\sigma_i - \sigma_j)$. $\Phi(\cdot)$ is the cumulative standard normal distribution, and σ_i and σ_j are location parameters of the items on the continuum of interest. Since $\pi(i,j) > .5$ when $\sigma_i > \sigma_j$, σ_i may be interpreted as the attractiveness of item i with respect to the trait of interest: the larger σ_i, the more preferable item i becomes with respect to the trait in question. Note that $\pi(i,j)$ is not a function of v: subjects are considered replications of each other. Since the derivation of the Case V variant of the Thurstone model is well-known (Thurstone, 1927; Torgerson, 1958, pp. 170–173), and since it is also presented extensively in Andrich (1978, pp. 452–455), it is not repeated here.

For practical purposes, the Thurstone model probability $\Phi(\sigma_i - \sigma_j)$ can be considered equivalent to the model probability $\Psi(\sigma_i - \sigma_j)$ as presented by Bradley and Terry (1952) and by Luce (1959); $\Psi(\cdot)$ denotes the cumulative logistic distribution function:

$$\frac{\exp(\cdot)}{1 + \exp(\cdot)} \qquad (1)$$

It is well-known that numerical differences between Φ and Ψ, when the argument of Ψ is appropriately transformed, are very small (Birnbaum, 1968, p. 399). The Thurstone and Bradley-Terry-Luce (BTL) models are analytically equivalent when doubly exponential distributions are assumed for the discriminal processes in the former model (Yellot, 1977).

Thus, following Andrich, in the following sections the ''Thurstone model'' refers to the model in which the pairwise choice probability $\pi(i,j)$ is specified as $\Psi(\sigma_i - \sigma_j)$. This model is to be ''equated'' in some way to the Rasch model.

The Rasch Model

In the Rasch model, the probability $\pi_v(i)$ of subject v responding positively to item i is assumed to be given by $\Psi(\xi_v^* - \sigma_i^*)$, in which ξ_v^* and σ_i^* are location parameters for the subject and the item respectively on the trait at issue (Fischer, 1974; Rasch, 1960). Since $\pi_v(i)$ is a decreasing function of σ_i^*, the larger σ_i^* the less preferable the item will be with respect to the trait in question, that is, the more it will tend to elicit a negative response from the subject. Since $\pi_v(i) > .5$ iff $\xi_v^* > \sigma_i^*$, the position of subject v on the trait in question may be conceived as representing the maximal position an item may have on this trait for the item being responded to more often positively than negatively. In comparison to the response process formalized by the Thurstone model, which assumes that items are compared pairwise and that the subject reports which item is judged to reflect more of the trait in question, the Rasch model can be described as assuming that the subject compares his/her position on the trait at issue with that of the item, and that a positive response is given whenever he/she perceives himself/herself to be positioned above the item.

The Rasch model has the characterizing property that the raw score $\sum_{i=1}^{K} a_{vi} = r_v$ is a sufficient statistic for ξ_v^* (cf. Mood, Graybill, & Boes, 1974, p. 326). This property makes it possible to use the procedure of conditional maximum likelihood (CML) estimation, in which the number of parameters is reduced by conditioning the likelihood of the data

on the sufficient statistics for the parameters to be eliminated. For example, the unknown ξ_i^* may be "conditioned away" by replacing them by the observable raw scores r_v, and subsequently the σ_i^* may be estimated independently of the subject sample.

Conditional estimation has both statistical and measurement theoretical advantages (Andersen, 1973a, 1973b; Gustafsson, 1980; Roskam & Jansen, 1984; van den Wollenberg, 1979, 1982); the latter have been labeled "specific objectivity" by Rasch (1966, 1977). Sample independence means that the item (subject) parameters are estimated independently of the specific positions of the subjects (items) on the trait at issue. The latter influence the parameter estimates only with respect to the precision of the estimators. (Andrich, 1978, pp. 458–460, devotes a section to sample independence. But a discussion of the merits of conditional estimation, and of its practicability, would be a digression. Therefore, its introduction here is motivated simply by pointing out that it is essential to Andrich's equivalency.) In Section 4, marginal maximum likelihood estimation, which has been presented as an alternative for CML, is briefly discussed with respect to the possibility of establishing a Thurstone-Rasch equivalency.

An Equivalency Between the Thurstone and Rasch Models

Since the response function $\pi(i,j)$ of the Thurstone model does not contain some sort of subject parameter and since the response function $\pi_v(i)$ of the Rasch model does, the parameter ξ_v^* of the latter has to be eliminated in some way to establish a formal equivalency between the two models. Clearly, the CML approach to estimation will serve this purpose. Note that in this way, sample independence has been chosen as the equivalency perspective. Whether the Thurstone model does indeed represent sample independent measurement is discussed in Section 4.

It is not difficult to derive that in the Rasch model, the probability $\pi_{ij}(1,0|a_{vi} + a_{vj} = 1) = \pi_{ij\cdot1}(1,0)$, i.e., that a subject v with raw score 1 on i and j will respond positively to item i ($a_{vi} = $

1) and negatively to j ($a_{vj} = 0$), is equal to $\Psi(\sigma_j^* - \sigma_i^*)$ (Andrich, 1978, pp. 457–458). To derive this result, local stochastic independence must be assumed:

$$\pi_{ij}(1,0|v) = \pi_v(i) \times [1 - \pi_v(j)] \quad , \tag{2}$$

which is an assumption commonly made in item response theory (IRT; cf. Lord, 1980, p. 19). The response probability $\Psi(\sigma_i - \sigma_j)$ of the Thurstone model and the response probability $\Psi(\sigma_j^* - \sigma_i^*)$ of the Rasch model are identical when it is assumed that $\sigma_i^* = -\sigma_i$. This assumption agrees, however, with the interpretation of σ_i in the Thurstone model: σ_i represents the degree to which item i is preferable with respect to the trait in question. In the Rasch model, σ_i represents the degree to which item i is not preferable with respect to the trait in question.

Until this point, the equivalency was motivated by the identity of two formulas. However, the mathematical model of Thurstone is something more than just a formula. It represents a substantive psychological theory about the processes involving paired comparisons. Consequently, an interpretation of the Rasch probability $\pi_{ij\cdot1}(1,0)$ in terms of the Thurstone model should encompass the psychological process the latter has been devised to represent.

Therefore, Andrich (1978, pp. 457–458) corroborated the formal Thurstone-Rasch equivalency by means of reasoning that is meant to demonstrate that there is also an equivalency on the level of assumed response processes underlying $\pi(i,j)$ in the Thurstone model and $\pi_{ij\cdot1}(1,0)$ in the Rasch model.

Interpreting $a_{vi} + a_{vj} = 1$ as

$$a_{vi} + a_{vj} = 1\text{: Subject } v \text{ perceives his/her}$$
$$\text{standing to be between the}$$
$$\text{two items } i \text{ and } j \tag{3}$$

$\pi_{ij\cdot1}(1,0)$ is conceived as the probability that v perceives j's position greater than his/hers (so that $a_{vj} = 0$) and item i's position is less (so that $a_{vi} = 1$). Then, using Interpretation 3, $\pi_{ij\cdot1}(1,0)$ is interpreted as representing a pair comparison process in which the comparison of i to j takes place by way of the three-element comparison of the latent trait position of v. In this way, the formal equivalency between the Thurstone and Rasch

models has been corroborated by an equivalency on the level of the item response theory embodied by these models.

2. Relationships Between the Thurstone and Rasch Models Considered from a Data-Theoretical Perspective

Another perspective for establishing a link between the Thurstone and Rasch models consists of the types of data for which these models are suited. In Section 3 Andrich's equivalency is reconsidered from the data-theoretical perspective introduced in this section. It then appears that the equivalency is based on an implicit assumption concerning the subject distribution that conflicts with the reasoning corroborating the equivalency (described in Section 1).

Types of Data

According to Coombs' (1964) categorization of data, the Thurstone model is especially suited for stimuli comparison data, and the Rasch model for single stimulus data ("direct response data" in Andrich's terminology). Stimuli comparison data consist of order relations on pairs of stimuli all representing elements from the same set, that is, of the Cartesian product $Q \times Q$, where Q denotes the stimulus set. An example is, "stimulus i is preferred to stimulus j in the pair (i,j)." Single stimulus data are $C \times Q$ data, that is, they consist of order relations on pairs of stimuli, the elements of which are taken from different sets (e.g., subjects and items). An example is, "subject v responds positively to item i." Clearly, the two types of data are incompatible because of the presence of the subject set C in single stimulus data.

Obviously, a possible data-theoretical bridge between $C \times Q$ data and $Q \times Q$ data may be $(C \times Q) \times (C \times Q)$ data, that is, preferential choice data. The latter consist of order relations between pairs of single stimulus data. An example is, "individual v prefers stimulus i more than individual v prefers stimulus j." Coombs (1964, p. 390ff) discussed a mapping of preferential choice data into stimuli comparison data by aggregating the $(C \times Q) \times (C \times Q)$ data over the elements of C, that

is, over the subjects. As is seen below, this aggregating over subjects can be specified as putting a special and restrictive requirement on the model that is preeminently suited for the analysis of preferential choice data, namely, Coombs' unfolding model. In this way, it is possible to write the Thurstone model as a restricted unfolding model (Coombs, 1964, p. 499ff). Finally, relating the unfolding model to the Rasch model, the Thurstone and Rasch models might be linked while the data-theoretical properties of both models are still taken into consideration. Note that in this way the "gap" between the two models is closed by the Coombs approach to item scaling.

Unidimensional Probabilistic Unfolding

A well-known technique for analyzing preferential choice data is Coombs' unfolding model (1964, chap. 5). According to the unfolding model, a subject v chooses i over j iff the distance δ_{vi} of the subject's scale value ξ_i^{**} to the item's scale value σ_i^{**} is less than the distance δ_{vj}:

$$\delta_{vi} < \delta_{vj} \quad (v \text{ prefers } i \text{ to } j) \quad . \tag{4}$$

Equation 4 formulates the deterministic unfolding model. Parametric probabilistic models of unfolding have been formulated by for example, Sixtl (1973) and Zinnes and Griggs (1974); Croon provided a recent overview (1983). For the present purpose this paper concentrates, however, on a model that has been presented by Jansen (1981, 1983). In this "Logistic Unfolding Model" (LUM) the probability of subject v preferring item i to item j with respect to the trait in question, is represented as:

$$\pi_v(i,j) = \Psi(\delta_{vj} - \delta_{vi}) \ (\sigma_i^{**} > \sigma_j^{**}) \quad . \tag{5}$$

The restriction on the item parameters is necessary to identify the model because different latent scales, namely, $\sigma_i^{**} < \sigma_j^{**}$ and $\sigma_i^{**} > \sigma_j^{**}$, are compatible with the same model probability $\pi_v(i,j)$, as is easily verified. Since the item's latent order on the unfolding dimension should be known beforehand to apply the LUM, the latter is especially suited for analyzing stimuli, the order of which either has already been determined from prior analyses or can be assumed on the basis of the content.

It is, however, also possible to determine "immediately" the probable item order from the preferential choice data at issue (Coombs, 1964, pp. 85–86).

According to Equation 5, $\pi_v(i,j) = .5$ iff $\xi_v^{**} = .5\,(\sigma_i^{**} + \sigma_j^{**}) = \sigma_{ij}^{**}$, that is, iff ξ_v^{**} is equal to the midpoint σ_{ij}^{**} of the item scale values. Furthermore, Equation 5 implies that

$$\delta_{vi} < \delta_{vj} \leftrightarrow \pi_v(i,j) > .5 \quad , \tag{6}$$

which may be conceived as the probabilistic counterpart of the unfolding rule formulated by Equation 4. Finally, $\pi_v(i,j)$ is an increasing function of ξ_v^{**} in the LUM when $\sigma_j^{**} < \xi_v^{**} < \sigma_i^{**}$, and $\pi_v(i,j)$ is constant when ξ_v^{**} is outside the interval $(\sigma_j^{**}, \sigma_i^{**})$. Figure 1, in which the LUM is depicted as a function of ξ_v^{**}, illustrates this.

Analogously to the Rasch scale, the LUM scale is a joint scale of both subjects and items. In the latter model, however, item i does not become more preferable with respect to the trait in question for all subjects when σ_i^{**} increases, because $\pi_v(i,j)$ is an increasing function of δ_{vi}, instead of σ_i^{**}. But, with subjects for whom

$$\xi_v^{**} > \sigma_i^{**}, \sigma_j^{**} \quad , \tag{7}$$

Equation 5 simplifies to $\Psi(\sigma_i^{**} - \sigma_j^{**})$, and in this case σ_i^{**} can be interpreted as the attractiveness, with respect to the trait in question, of item i. In the sequel, the σ_i^{**} scale in the LUM is interpreted as representing the degree to which items are preferable with respect to the trait in question according to those subjects whose trait positions are above all item positions.

Unilaterality

When all subjects are on the same side of the items on the latent scale, the item set is called *unilateral*. Unilaterality was introduced by Coombs (1964, p. 107ff) in the context of stochastic transitivity of pairwise choice probabilities, but subsequently he used it (p. 499ff) to relate the Thurstone model to stochastic unfolding. Coombs derived that the Thurstone model can be conceived as a model of probabilistic unfolding, in which the subject group is restricted to those subjects with respect to whom the stimuli are unilateral on the unfolding scale. Application of the Thurstone model to a population of subjects with varying scale values on an unfolding dimension requires unilaterality of the stimulus set, that is, the application would have to be limited to a specific subgroup of subjects. In such a case, the subjects may be effectively replications, that is, they need not necessarily be represented in the model of measurement.

Figure 1

Example of the Item Characteristic Curve in the
Logistic Unfolding Model Formulated in Equation 5
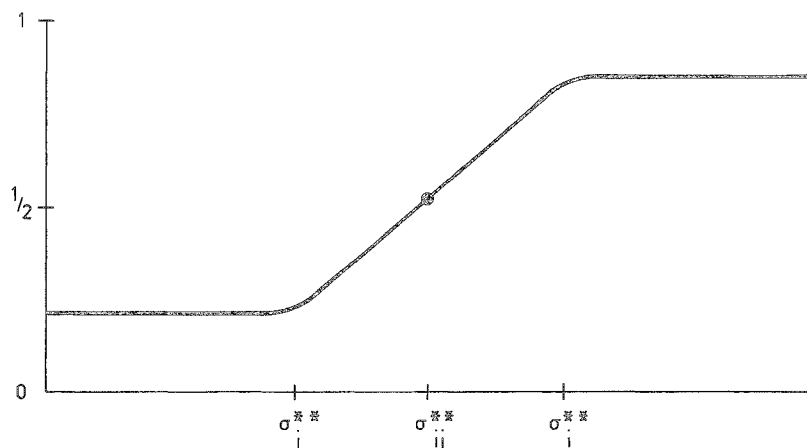
These findings can be easily transferred to parametric probabilistic unfolding and the Thurstone model $\Psi(\sigma_i - \sigma_j)$. As in the case of unilaterality formulated by Equation 7, the LUM reduces to the Thurstone model; the unfolding counterpart of the latter model is the unilateral LUM. Note that the interpretation of σ_i is the same in both models: the degree to which items are preferable with respect to the trait in question according to those subjects who constitute the reference population of the models. Thus, the unfolding counterpart of the Thurstone formulation $\Psi(\sigma_i - \sigma_j)$ is the unilateral LUM. Analogously, the unfolding counterpart of the original Thurstone formulation $\Phi(\sigma_i - \sigma_j)$ is the model proposed by Zinnes and Griggs (1974), since in the latter model $\pi_v(i,j) \rightarrow \Phi(\sigma_i - \sigma_j)$ when $\xi_v \rightarrow +\infty$.

It should be noted here that there exist parametric probabilistic unfolding models in which ξ_v is not eliminated for unilateral stimuli; an example is given in Section 4.

Conclusion

The Thurstone model is identical to a special case (unilaterality) of a specific form (LUM) of probabilistic unfolding. This reinterpretation of the Thurstone model does not require additional assumptions. With respect to the complete LUM however, the subject sample has been limited to those subjects who comply with Equation 7: The reference population of the Thurstone model is a subset of the reference population of the LUM. In the next section it is investigated whether this restriction of the subject population by the Thurstone model has implications for the relationships between the Thurstone and Rasch models as established in Andrich's (1978) analysis.

3. Andrich's Equivalency Considered from a Data-Theoretical Perspective

When Andrich's reasoning holds, it follows from Sections 1 and 2 that

$$-\sigma_i^* \text{ (Rasch)} = \sigma_i \text{ (Thurstone)}$$
$$= \sigma_i^{**} \text{ (unilateral LUM)} , \qquad (8)$$

except for arbitrary choices of the origins and units of the scales. Note that the equality between the Rasch and Thurstone parameters may be empirically invalid, but that the equality between the Thurstone and unilateral LUM parameters represents a formal identity that always holds.

Equation 8 expresses that at the level of the item parameters, there exists a triple equivalency between the Thurstone model, the Rasch model, and the unilateral LUM. However, what about the *subject* parameters of the latter two models? Below, five cases are treated. First, it may be assumed that ξ_v^* (Rasch) $\neq \xi_v^{**}$ (unilateral LUM) (Section 3). Second, it may be assumed that $\xi_v^* = \xi_v^{**}$, that is, both models refer to exactly the same subject density (Section 3). Third, it may be asserted that because of sample independence, the (non)equivalency between ξ_v^* and ξ_v^{**} is irrelevant (Section 4). In such a case, statistical elimination of the subject in the Rasch model (by CML) is considered equivalent to "experimental" elimination of the subject in the Thurstone model/unilateral LUM. Fourth, it may be argued that the latter assertion is not correct and that in both the models which are to be "equated," sample independence is to be conceived in the same way, namely, as statistical elimination of the subject parameter (Section 4). And fifth, it might be investigated in what way the marginal maximum likelihood approach to estimation in the Rasch model eliminates ξ in the same way as in the Thurstone model (Section 4).

Nonequivalent Subject Parameters

When it is assumed that $\xi_v^* \neq \xi_v^{**}$, the Rasch model and the unilateral LUM are not equivalent with respect to subject measurement. However, since the unilateral LUM and the Thurstone model are identical, this would imply that the Thurstone-Rasch equivalency would only refer to the items, and as a consequence would be of little value for applied measurement.

Equivalent Subject Parameters

Suppose it is assumed that

$$\xi_v^* \text{ (Rasch)} = \xi_v^{**} \text{ (unilateral LUM)} . \qquad (9)$$

This would imply that the property of unilaterality of the LUM is transferred to the Rasch model, and that as a consequence the subject parameters in the latter model would have to be assumed to be all located above the item parameters. But, this conflicts with the fact that the Rasch model does not impose any restrictions on the trait locations of the subjects from the reference population. Thus, though Equation 9 may lead to an equivalency between the unilateral LUM and the Rasch model as instruments for subject measurement, the equivalency would be obtained by restricting the reference population of the Rasch model in a nontrivial way.

When Assumption 9 would be accepted, the applicability of the Rasch model would depend on the unilaterality of the items, and consequently item-independent subject measurement (and subject-independent item measurement) would be lost. A way to "save" the latter would be to define the subject population of the Rasch model in terms of the subject population of the Thurstone model/unilateral LUM, so that for all samples from this restricted population, items would be unilateral. In this new subject population, discrimination of the items would be low. On account of their responses, subjects would be effectively indistinguishable, that is, for practical purposes they might as well be considered replications. This is the very reason why, for this particular subject group, the Rasch model appears empirically equivalent to the Thurstone model.

Thus, considering Andrich's equivalency from a data-theoretical perspective yields that it is based on the assumption that the Rasch model is restricted to the subset of subjects who constitute the Thurstone reference population. The Thurstone and Rasch models become empirically equivalent when subject influences are de facto absent in the Rasch model.

Unilaterality of the subject set should be considered a legitimate assumption, the test of which is in the data, in order to achieve a Thurstone-Rasch equivalency. However, it seems to contradict the reasoning corroborating the equivalency. For, according to Equation 3, subjects are assumed *between* the items instead of above them on the Rasch scale. Interpreting the Rasch probability $\pi_{ij,1}(1,0)$ in terms of some sort of pair comparison process with the subject perceiving his/her standing between i and j runs counter to the unilaterality assumption, which is necessary to achieve an equivalency between the Thurstone and Rasch models as measurement models. In the Discussion (Section 5) an alternative, more plausible interpretation of the Rasch probability $\pi_{ij,1}(1,0)$ is presented.

4. Sample Independence as an Equivalency Aspect

It might be argued that the equivalency between the Thurstone model, or unilateral LUM, and the Rasch model holds only for the items, and that the relation between ξ^* and ξ^{**} is irrelevant because item measurement is independent of the subject sample in both approaches to item scaling. According to this argument, statistical elimination of the subject parameter in the Rasch model by means of conditioning on sufficient statistics is equivalent to "experimental" elimination of the subject in the Thurstone model by restricting the reference population to those subjects with respect to whom all items would be unilateral. This seemed to be Andrich's (1978) position:

> However, a critical point in unifying the approaches is that this [subject; *pj*] parameter, which can be eliminated explicitly by statistical conditioning in Rasch's model, is also eliminated in the paired comparison design. (p. 452)

This position defines a perspective for deriving a pair comparison counterpart of the Rasch model. The question is, however, whether the Thurstone model is indeed the sample-independent pair comparison model sought for. In the following sections this question is studied, an alternative sample-independent pair comparison counterpart of the Rasch model is presented, and finally, it is briefly investigated whether marginal maximum likelihood estimation may establish a Thurstone-Rasch equivalency.

Sample Independence and the Thurstone Model

With the Rasch model, estimation of the item parameters is independent of the subject parameter because the raw score r_v is sufficient for the latter. However, since r_v is a perfect substitute for the subject parameter with respect to statistical information, the conditional likelihood of the data given the raw score still is a function of v, that is, individual differences are still present in the conditional Rasch model. Thus, by conditioning the item response probability on r_v, the subject population is not restricted in any way, contrary to the unilateral LUM.

The foregoing implies that, with respect to sample independence in its proper sense of conditional estimation, a natural "preferential choice" Thurstonian counterpart of the Rasch model would be a pairwise choice model that both allows for individual differences and possesses sufficient statistics for the individual parameters. As the Thurstone model $\Phi(\sigma_i - \sigma_j)$ and the BTL model $\Psi(\sigma_i - \sigma_j)$ have been treated as equivalent in this paper, the desired Thurstonian model would have to be a BTL model on the individual level, that is, it would have to be of the form $\Psi(\sigma_{vi} - \sigma_{vj})$. As a consequence, the response probability $\pi_v(i,j)$ would have to comply with the "individual constant ratio rule,"

$$\pi_v(i,j) = \frac{\pi_v(i)}{\pi_v(i) + \pi_v(j)} \quad , \qquad (10)$$

that is characteristic for a BTL model (Suppes & Zinnes, 1963, p. 49).

But Andersen (1976, p. 155) has proven that there does not exist a BTL model $\Psi(\sigma_{vi} - \sigma_{vj})$ on the individual level that possesses sufficient statistics for the individual parameters. This implies, for instance, that the Rasch model cannot be written as a BTL model on the individual level, which can be demonstrated as follows.

The probability $\pi_{ij,1}$ $(1,0)$ in the Rasch model can be written as the function

$$\frac{\pi_v(i)}{\pi_v(i) + \pi_v(j) \left[\dfrac{1 - \pi_v(i)}{1 - \pi_v(j)} \right]} \quad , \qquad (11)$$

of the single stimulus probabilities $\pi_v(i)$. Since Equations 10 and 11 are identical iff $\pi_v(i) = \pi_v(j)$, that is, iff $\sigma_i^* = \sigma_j^*$, the transformation represented by Equation 11 will conflict in every nontrivial situation with the ratio rule formulated by Equation 10. As a consequence, the Rasch model, even transformed according to Equation 11, cannot be written as a BTL model on the individual level.

A Sample-Independent Pair Comparison Model

It is possible to conceive models for preferential choice data that both allow for individual differences and possess sufficient statistics for the subject parameters, but Andersen's (1976) result shows that these necessarily will not be individual BTL models. An example is the model of probabilistic unfolding that was proposed by Sixtl (1973) and elaborated by Jansen (1981, 1983). In this "Rasch Homogeneous Unfolding Model" (RHUM), the probability of subject v preferring item i above item j with respect to the trait at issue is represented as:

$$\pi_v(i,j) = \Psi(\xi_v^{***} - \sigma_{ij}^{***})(\sigma_i^{***} > \sigma_j^{***}), \qquad (12)$$

where again $\sigma_{ij}^{***} = .5 \ (\sigma_i^{***} + \sigma_j^{***})$. When $\sigma_i^{***} = \sigma_j^{***}$, $\pi_v(i, j)$ is defined to be equal to .5, as Figure 2 shows.

Scaling by the RHUM works as follows. K stimuli, the order of which on the latent scale of interest is known, are presented in pairs (i,j), with $\sigma_i^{***} > \sigma_j^{***}$, to N subjects, asking for a preference response, for example, "Choose that stimulus that represents to you the most of the trait in question." The responses are collected in a $N \times \binom{K}{2}$ data matrix A with dichotomous entries a_{vij}. When $a_{vij} = 1$, the first stimulus of the pair (i,j) was chosen by v, and when $a_{vij} = 0$, the second stimulus was chosen. To estimate the item parameters directly and independently of the sample of subjects, instead of using the midpoint estimates as obtained with a standard Rasch analysis on the data A, the procedure of the "linear logistic test model" can be applied (cf. Fischer, 1974, 1983), as was demonstrated by Jansen (1981, 1983).

Figure 2
Example of the Item Characteristic Curve in the
Rasch Homogeneous Unfolding Model Formulated in Equation 12
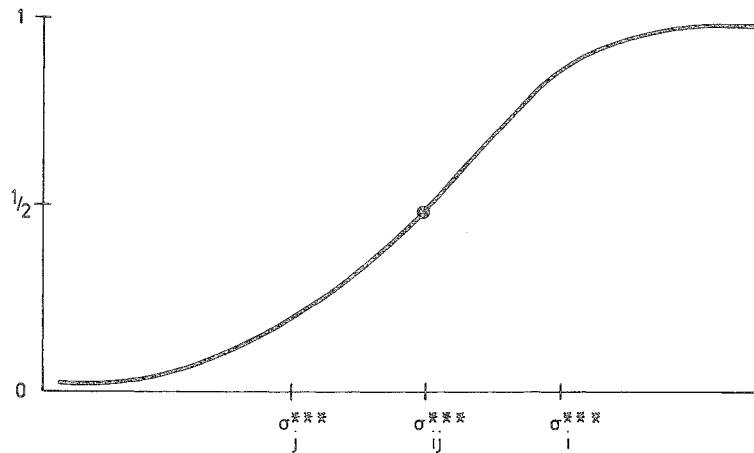
Thus, with respect to conditional estimation, the RHUM represents a preferential choice counterpart of the Rasch model. However, the RHUM is not a BTL model, since for instance, ξ^{***} is not canceled in $\pi_v(i,j)$ when both i and j are unilateral to v (as may be easily verified in Equation 12). Therefore, though the RHUM may be considered a counterpart of the Rasch model in the case of preferential choice data, it cannot be written as a Thurstone model. It would, of course, be interesting to study the relationships between the Rasch model and the RHUM, for example, conditions under which they are equivalent with respect to applied measurement, but the Thurstone model would not play a part in such a comparison.

Marginal Maximum Likelihood Estimation

Until now, the subject parameter was eliminated from the Rasch model by means of the CML procedure. Recently, however, the method of marginal maximum likelihood (MML) has been presented as an alternative for CML (Bock & Aitkin, 1981; Thissen, 1982). In CML, response probabilities are independent of the subject parameter at the level of an individual subject (by conditioning on that subject's raw score), whereas in MML such independence is obtained at the level of an "average

subject" (by integrating over ξ), that is, at the level of a subject sampled randomly from the reference population. The latter subject is also the frame of reference for the Thurstone model, so the MML approach to estimation could be a base for comparing the Thurstone and Rasch models.

However, applying MML to the individual pair comparison probability as defined in Equation 2 does not generally result in the Thurstone model $\Psi(\sigma_i - \sigma_j)$. Only in some ad hoc way, by choosing an appropriate subject distribution, might MML establish an equivalency between the Thurstone and Rasch models. Still, this equivalency would be valid for an "average" subject only. Thus, eventually the comment of Section 3 would apply here as well.

5. Discussion

The data-theoretical perspective revealed that a Thurstone-Rasch equivalency is obtained in the case of unilaterality of the subject with respect to the items on the latent scale. This implies that an explanation of empirical equivalencies between Thurstone and Rasch analyses (as e.g., observed in the study of Wainer, Fairbank, & Hough, 1978) should be sought, with respect to the Rasch model, in an incidental unilaterality of the item set. For example, items may be too "easy" or too "difficult"

for the sample, or the range of subject abilities may be low so that subjects become effectively replications.

Whether the condition of unilaterality holds in specific applications is an empirical matter. However, at a theoretical level it prevents a pair comparison interpretation of the Rasch probability $\pi_{ij\cdot 1}$ (1,0) in which the subject is assumed to be between the items. As Roskam and Jansen (1984) demonstrated (cf. also Jansen, 1983, Roskam, 1983), this probability can be conceived as defining a sample independent stochastic ordering of the items i,j. In this interpretation, the position of the subject is irrelevant, and, furthermore, no direct or indirect intrapersonal comparison of the items is assumed. The stochastic item ordering is derived from single stimulus data assuming Ducamp and Falmagne's (1969) axiom of composite transitivity.

As the previous sections demonstrate, comparisons between scaling models are important. Item response models have proliferated and it seems worthwhile, both for practical and for theoretical purposes, to investigate the relationships between various approaches to scaling. Connections between models may be studied both theoretically (i.e. purely mathematically) as was the case in Andrich's study (1978) and in this paper, or empirically, as was, for instance, the case in Wainer et al. (1978).

Theoretical studies have practical implications when they are aimed at detecting generally interpretable and simply (i.e. globally) testable conditions for the validity of measurement models. These applicability requirements should be phrased in terms of characteristics of the data, but in a more specific way than in Coombs' (1964) typology. It is believed that situationally specific practicability of scaling models can be better judged with respect to a few general and easily observable data conditions derived from them, than by trying to fit, in an inductive trial-and-error way, the best-fitting model with every set of observations. Apart from Andrich's (1978) study and the present paper, an example of such a theoretical model analysis for practical purposes is Jansen's (1981, 1983) study of the suitability of the RHUM for representing pair comparison data.

In the same manner the Thurstone/BTL model could be contrasted theoretically to the RHUM. Obviously, the former model does not allow for individual differences, whereas the latter does (this is a classification category in Coombs', 1964, theory of data). More specifically: the former predicts strong stochastic transitivity of pairwise preferential choices for all kinds of stimulus triples (Suppes & Zinnes, 1963, p. 59); the latter predicts moderate or strong stochastic transitivity depending on the specific stimulus triple kinds.

Apart from gaining a better insight into the suitability of specific measurement models for specific data types, theoretical model comparisons are of importance for assessing the validity of measurement scales. If different methods of collecting item observations (direct responses and paired comparisons) and corresponding different models of measurement (Rasch's and Thurstone's or Coombs') yield essentially the same results, then the validity of both sets of measurements has been strongly corroborated. Especially in the field of attitude measurement, such a multimethod approach to measurement may be an important step toward separating the attitude construct from the present day's specific and seemingly incompatible types of measurement procedures and corresponding analysis models (e.g., Likert scaling, semantic differential, magnitude estimation, or Thurstone scaling; cf. Fishbein & Azjen, 1972). In this way, both attitudinal measurement and IRT would be supported.

References

Andersen, E. B. (1973a). *Conditional inference and models for measuring*. Kopenhagen: Mental-hygiejnisk Forlag.

Andersen, E. B. (1973b). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123–140.

Andersen, E. B. (1976). Paired comparisons with individual differences. *Psychometrika, 41,* 141–157.

Andrich, D. (1978). Relationships between the Thurstone and the Rasch approaches to item scaling. *Applied Psychological Measurement, 2,* 451–462.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 315–479). Reading MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An appli-

cation of an EM algorithm. *Psychometrika, 46,* 443–459.

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

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Croon, M. (1983). *Statistical unfolding models* (Interim Report). Tilburg, The Netherlands: Faculteit Sociale Wetenschappen, Katholieke Hogeschool Tilburg.

Ducamp, A., & Falmagne, J. C. (1969). Composite measurement. *Journal of Mathematical Psychology, 6,* 359–390.

Fischer, G. H. (1974). *Einführung in die theorie psychologischer tests.* Bern: Huber.

Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48,* 3–26.

Fishbein, M., & Azjen, I. (1975). *Belief, attitude, intention and behavior. An introduction to theory and research.* New York: Addison-Wesley.

Gustafsson, J.-E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 33,* 205–233.

Jansen, P. G. W. (1981). Spezifisch objektive Messung im Falle nicht-monotoner Einstellungsitems. *Zeitschrift für Sozialpsychologie, 12,* 169–185.

Jansen, P. G. W. (1983). *Rasch analysis of attitudinal data.* Doctoral Dissertation, Rijks Psychologische Dienst, The Hague, The Netherlands.

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

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

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). Tokyo: McGraw-Hill.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1966). An informal report on a theory of objectivity in comparisons. In L. J. van der Kamp & C. A.J. Vlek (Eds.), *Proceedings of the NUFFIC International Summer Session in Science at "Het Oude Hof"* (pp. 1–19). The Hague, The Netherlands.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy, 14,* 58–94.

Roskam, E. E. (1983). Allgemeine Datentheorie. In H. Feger & J. Bredenkamp (Eds.), *Messen und testen* (Forschungsmethoden der Psychologie; Enzyklopädie der Psychologie, Serie I, Part I, pp. 1–135). Gottingen: Hogrefe.

Roskam, E. E., & Jansen, P. G. W. (1984). A new derivation of the Rasch model. In E. Degreef & J. Van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 293–307). Amsterdam: Elsevier Science Publishers B. V. (North-Holland).

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

Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology,* (Vol. I, pp. 1–76). New York: Wiley.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47,* 175–186.

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

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

van den Wollenberg, A. L. (1979). *The Rasch model and time-limit tests.* Doctoral Dissertation, Studentenpers, University of Nijmegen, The Netherlands.

van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47,* 123–140.

Wainer, H., Fairbank, D. T., & Hough, R. L. (1978). Predicting the impact of simple and compound life change events. *Applied Psychological Measurement, 2,* 313–322.

Yellot, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment and the double exponential function. *Journal of Mathematical Psychology, 15,* 109–144.

Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic multidimensional unfolding analysis. *Psychometrika, 39,* 350–372.

Author's Address

Send requests for reprints or further information to Paul G. W. Jansen, Afdeling Bedrijfspsychologie, Centrale Directie PTT, P.O. Box 30000, 2500 GA The Hague, The Netherlands.