Multidimensional Rasch Models for Partial-Credit Scoring
Henk Kelderman
Vrije Universiteit, Amsterdam

Rasch models for partial-credit scoring are discussed and a multidimensional version of the model is formulated. A model may be specified in which consecutive item responses depend on an underlying latent trait. In the multidimensional partial-credit model, different responses may be explained by different latent traits. Data from van Kuyk's (1988) size concept test and the Raven Progressive Matrices test were analyzed. Maximum likelihood estimation and goodness-of-fit testing are discussed and applied to these datasets. Goodness-of-fit statistics show that for both tests, multidimensional partial-credit models were more appropriate than the unidimensional partial-credit model. Index terms: \( \chi^2 \) testing, exponential family model, multidimensional item response theory, multidimensional Rasch model, partial-credit models, Progressive Matrices test, Rasch model.

Responses to educational and psychological test questions can be scored partially correct rather than simply correct or incorrect. To relate a person's responses to the person's underlying latent trait, item response models for dichotomously scored data (Birnbaum, 1968; Lord, 1980; Lord & Novick, 1968; Rasch, 1960/80) have been generalized to polytomous ordered data (Andrich, 1978a, 1978b; Glas & Verhelst, 1989; Masters, 1982; Muraki, 1990; Rost, 1988; Samejima, 1969). In some applications, however, it is questionable whether fully correct answers require the same trait as partially correct answers. A Rasch-type multidimensional partial-credit model (PCM) was formulated for data in which different answers depend on different traits.

**Partial-Credit Models**

A Unidimensional Model

For the unidimensional PCM (PCM; Andrich, 1978a; Masters, 1982), let \( \pi_{ij} \) be the probability of response \( x \) of person \( j \) to item \( i \) and \( r_i \) be the number of possible (partial) credit responses for item \( i \). The PCM in terms of the log-odds of consecutive item responses can be written as

\[
\Omega_{ij} = \log \left( \frac{\pi_{ij}}{\pi_{ij+1}} \right) = \theta_j - \delta_{x} \quad j = 1, \ldots, N, \quad i = 1, \ldots, n, \quad x = 1, \ldots, r_i.
\]

with

\[
\sum_j \sum_x \delta_{x} = 0
\]

to fix the scale, and log denotes the logarithm to the base \( e = 2.718281 \). That is, the log-odds of consecutive responses can be written as the difference of a person parameter \( \theta_j \) and a response parameter \( \delta_x \). The person parameter describes the trait level of the person. As the person parameter becomes higher, the log-odds of giving a higher quality response \( x \) rather than a lower quality response \( x - 1 \) becomes higher. The threshold parameter describes the difficulty of the item response. As the threshold parameter becomes higher, the log-odds of giving the response \( x \) rather than the response \( x - 1 \) becomes lower. A similar model for the log-odds of nonconsecutive responses is
This model is a generalization of the Rasch model for dichotomously scored \((r_1 = 1)\) items. It has the useful property that objects—items or persons—can be compared independently of other objects within the frame of reference (Rasch, 1961, 1977). This property is called specific objectivity (Fischer, 1995; Masters, 1987). In Equation 1, specific objectivity means that two different persons, say \(j\) and \(j'\), can be compared by

\[
\Omega_{\theta j} - \Omega_{\theta j'} = \theta_j - \theta_{j'}
\]

independently of the response \(x\) to item \(i\) \((i, x)\). Two different responses [e.g., \((i, x)\) and \((i', x')\)] can be compared by

\[
\Omega_{\delta x} - \Omega_{\delta x'} = \delta_{x'} - \delta_x
\]

regardless of the person \(j\).

Other unique measurement properties of the Rasch model are that the responses of all items have identical (loglinear) interactions with all other variables in the nomological network (Kelderman, 1995) and that the items follow composite transitivity (Roskam & Jansen, 1984).

Rasch models describe the trait \((\theta)\) and difficulty \((\delta)\) parameters of individual persons and items, respectively (Rasch, 1960). Not considered in these models is the statistical generalization to other persons from the same population of intended respondents, nor to other items from the same universe of intended items. Therefore, no sampling distributions are specified for person or item parameters. See Molenaar (1995) for a discussion of the different sources of randomness in item response theory (IRT) models.

### Multidimensional Partial-Credit Models

The 1PCM with trait parameter \(\theta_j\) can be generalized to a multidimensional PCM (MPCM) with \(s\) trait parameters \(\theta_q\) \((q = 1, \ldots, s)\). Let \(w_{qix}\) be an indicator variable that takes the value 1 if the log-odds \(\Omega_{\theta q}\) of giving response \(x\) rather than \(x - 1\) depends on latent trait \(\theta_q\) and 0 otherwise. The MPCM is then

\[
\Omega_{\theta q} = \sum_{q=1}^{s} w_{qix} \theta_q - \delta_x.
\]

In this model, the person’s ability to give a response to a particular item is considered to be equal to a weighted sum of more basic trait parameters. In the general case of this MPCM, the weights \(w_{qix}\) may take discrete non-negative values \((0, 1, 2, \ldots)\).

Like the 1PCM, this MPCM has the useful property that response alternatives can be compared independently of the person, and persons can be compared independently of the alternatives, provided that both responses depend on the trait of interest.

The model can be written in terms of the response probabilities \(p_{qx}\) rather than the log-odds \(\Omega_{\theta q}\). Next, the probability of a response pattern and of the entire data matrix are provided.

From Equation 1,

\[
\Pi_{\theta q} = \exp \left( \sum_{y=1}^{r} \Omega_{\theta qy} \right) \Pi_{\theta 0}, \quad x = 1, \ldots, r.
\]

Furthermore from

\[
1 = \sum_{x=0}^{r} \Pi_{\theta x} = \Pi_{\theta 0} + \sum_{x=1}^{r} \Pi_{\theta x} = \Pi_{\theta 0} + \sum_{x=1}^{r} \exp \left( \sum_{y=1}^{r} \Omega_{\theta qy} \right),
\]
\[ \Pi_{ij} = \frac{1}{1 + \sum_{x=1}^{r} \exp \left( \sum_{y=1}^{s} \Omega_{xy} \right)} , \] (9)

so that from Equations 6–9

\[ \Pi_{ij} = \frac{\exp \left[ \sum_{y=1}^{s} \left( \sum_{q=1}^{r} w_{qy} \theta_{qy} - \delta_{iy} \right) \right]}{1 + \sum_{x=1}^{r} \exp \left[ \sum_{y=1}^{s} \left( \sum_{q=1}^{r} w_{qy} \theta_{qy} - \delta_{iy} \right) \right]} , \] (10)

where

\[ \sum_{y=1}^{s} (\cdot) = 0. \] (11)

Note that if \( s = 1 \) and \( w_{qy} = 1 \), the IPCM results.

If the consecutive odds interpretation of the model is not appropriate for the data, an alternative parameterization of Equation 10 is

\[ \Pi_{ij} = \frac{\exp \left( \sum_{y=1}^{s} B_{qy} \theta_{qy} + \phi_{ix} \right)}{1 + \sum_{y=1}^{s} \exp \left( \sum_{q=1}^{r} B_{qy} \theta_{qy} + \phi_{iq} \right)} , \] (12)

where

\[ \phi_{ix} = 0 , \] (13)

\[ \phi_{ix} = \sum_{y=1}^{s} \delta_{iy} \ (x = 1, \ldots, r_i) , \] (14)

and

\[ B_{qy} = 0 , \] (15)

\[ B_{qy} = \sum_{y=1}^{s} w_{qy} \ (x = 1, \ldots, r_i) , \] (16)

and conversely,

\[ \delta_{iy} = \phi_{i(x-1)} - \phi_{ix} \ (x = 1, \ldots, r_i) , \] (17)

and

\[ w_{qy} = B_{qy} - B_{q(x-1)} \ (x = 1, \ldots, r_i) . \] (18)

This model is a generalization of Rasch's original multidimensional model (Andersen, 1973b; discussed below); it also is a confirmatory multidimensional IRT model.

The model is multidimensional because the \( \theta_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jr}) \) are vector valued so that the trait of each person can be represented by a point in an \( r \)-dimensional real latent space (Lord & Novick, 1968, p. 359). Note that this multidimensional Rasch model (the MPCM) does not specify a population distribution for \( \theta_j \).
because the model is concerned with individual characteristics rather than population characteristics.

The model is confirmatory because the \( B \) weights, indicating the strength of the relation between the item response and \( \theta_j \), must be specified. The model describes the relationship between \( \theta_j \) and the probability of each of the item responses. If a person’s value on \( \theta_j \) becomes higher, the probability of responses with larger \( B \) weights on that trait become higher. Conversely, if a person has many item responses that have high \( B \) weights for a certain latent trait, the model states that these responses have a high probability under a high value of this latent trait, so that these responses indicate a high value of the trait.

To specify an ordinary PCM, \( w_{qij} = 1 \) (\( q = 1; i = 1, \ldots, n; x = 0, \ldots, r \)), which is equivalent to specifying \( B_{qij} = x, (q = 1; i = 1, \ldots, n; x = 0, \ldots, r) \). Figure 1a describes the influence in the PCM of the latent trait \( \theta_j \) on each of the responses. The \( B \) weights over the arrows describe the strength of the relation. In this model, the correct response (\( x = 2 \)) depends 2 times more on \( \theta_j \) than the partially correct response (\( x = 1 \)). Figure 1b shows the same situation for a 2PCM. The partially correct response and the fully correct response both depend on \( \theta_j \). In addition, the fully correct response also depends on \( \theta_2 \).

![Figure 1 Models for Three-Category Items](image)

Estimating the Multidimensional Partial-Credit Model

*The likelihood of the data under the MPCM.* Denote the response \( x \) of person \( j \) to item \( i \) as \( x_{ij} \) and let \( x_j = (x_{1j}, \ldots, x_{nj}) \) be the vector of responses of person \( j \). Let

\[
t_j = \sum_{i=1}^n \sum_{y=1}^{r_j} w_{qy} \left( = \sum_{i} B_{qiy} \right)
\]

be person \( j \)'s sum of weights on trait \( q \); \( t_j = (t_{j1}, \ldots, t_{jn}) \) be the vector of weight-sums; and \( \delta \) be the vector of threshold parameters. If it is assumed that the person's responses are independent, the probability of this response vector can be derived from Equations 6 and 7 as

\[
p(x_j | \theta) = \prod_{i=1}^n \prod_{q=1}^{r_i} \exp \left[ \sum_{y=1}^{r_y} w_{qy} \theta_q - \delta_y \right] + \sum_{i=1}^n \ln \Pi_{q=1}^{r_i} = g(\delta, \theta, t_j) h(\delta, x_j),
\]

where

\[
g(\delta, \theta, t_j) = \exp \left[ \sum_{q=1}^n \sum_{y=1}^{r_q} \theta_q t_{qy} \right] + \sum_{r=1}^{n} \ln \Pi_{y=1}^{r} 
\]

and

\[
h(\delta, x_j) = \exp \left[ \sum_{i=1}^n \sum_{y=1}^{r_i} \delta_y \right].
\]
Estimation of parameters. The model in Equation 20 contains \( n \) item parameters and \( N \) person parameters. It may seem natural to estimate both item parameters and person parameters maximizing the likelihood

\[
P(X) = \prod_{j=1}^{N} p(x_j | \theta_j),
\]

where \( X = (x_1, ..., x_n)' \) is the data matrix. However, as the number of person parameters increases with \( N \), the parameter estimates become inconsistent (Neyman & Scott, 1948). To remedy this situation, a conditional estimation procedure is used.

The model in Equation 20 is an exponential family distribution (Lehman, 1983, p. 26), and \( t_j = (t_{j1}, ..., t_{jp}) \) is a sufficient statistic for the person parameters \( \theta_j = (\theta_{j1}, \theta_{j2}, ..., \theta_{jp}) \). This implies that all statistical information in the data about the person’s position in the multidimensional latent space \( \theta_j \) is contained in the \( t_j \); that is, there is a one-to-one relation between parameter estimates and the statistic. Note that \( t_j \) is a statistic because the model is confirmatory; that is, it depends on the data through weights \( (B_{eq} \text{ or } w_{eq}) \) that are not estimated but specified by the user. Exponential family models allow for conditional inference (Lehman, 1983) as follows.

If \( \Sigma y_{ij} \) is defined as the sum over all possible response vectors \( y \) that give rise to a weight-sum vector \( t_j \), the probability of a vector \( t_j \) can be written as

\[
P(t_j | \theta) = \sum_{y} p(y | \theta) = g(\delta, \theta, t_j) \sum_{\eta_j} h(\delta, y).
\]

A convenient property of Rasch models is that the conditional probability of \( x_j \) given \( t_j \) does not depend on \( \theta_j \):

\[
P(x_j | t_j) = \frac{P(x_j | \theta_j)}{P(t_j | \theta_j)} = h(\delta, x_j) / \sum_{\eta_j} h(\delta, y) = h(\delta, x_j) / f(\delta, t_j),
\]

which results from division of Equation 20 by Equation 24. This means that estimation of the threshold parameters \( \delta \) can be pursued independently of the estimation of person parameters. Note that this property holds for both multidimensional and unidimensional models, because Equation 20 is an exponential family model.

Let \( S \) be the number of persons with weight-sum vector \( t \) and let \( R \) be the number of persons with a response of at least \( y \) to item \( i \). Furthermore, let \( T = (t_{1}, ..., t_{p})' \) be the matrix of weight-sums. The conditional probability of the data given the weight-sums then becomes

\[
P(X | T) = \prod_{j=1}^{N} p(x_j | t_j) = \exp \left[ \sum_{y} \log h(\delta, x_j) + \sum_{\eta_j} \log f(\delta, t_j) \right] = \exp \left[ \sum_{i} \delta_i \sum_{y} R_{iy} - \sum_{i} \tau_i S_i \right],
\]

where \( \tau_i = \log f(\delta, t) \) is a constant of proportionality. Viewed as a function of the parameters \( \delta \), this is the conditional likelihood function, which takes its maximum for the maximum likelihood estimator of \( \delta \). This likelihood has the form of an exponential family with sufficient statistics \( R \) for the threshold parameters \( \delta_i \) and numbers \( S_i \) for each of the proportionality constants \( \tau_i \) (Lehman, 1983, p. 43). Standard theory for exponential family models yields the likelihood equations \( R_i = E(R_{iy}) \) and \( S_i = E(S_i) \). Solving these equations for the parameters \( \delta \) and \( \tau \) gives the maximum likelihood estimates (MLEs) of the parameters. The equations, however, have no direct solution so they must be solved by iterative methods. For the case of the PCM (\( s = 1 \)), efficient algorithms have been described by Andersen (1973a, 1977) and Fischer (1974). For the case of a discrete-data-exponential-family likelihood of the form of Equation 26, more general algorithms for the analysis of discrete data (Andersen, 1980; Bishop, Fienberg, & Holland, 1975; Haberman, 1979; Kelderman, 1992) may be used. Appendix A describes an adaptation of a Deming-Stephan (Deming & Stephan, 1940; Bishop et al., 1975, p. 84) algorithm; Appendix B describes the Newton-Raphson algorithm. Both algorithms yield MLEs.
Newton-Raphson is faster, but it may break down if the starting values are far from the MLEs. It also requires the computation of second-order derivatives of the likelihood function. The Deming-Stephan algorithm, however, is not sensitive to the choice of starting values and requires the sufficient statistics $R_n$ and $S_n$ only. Deming-Stephan takes more iterations than Newton-Raphson, but each iteration takes less computation. The Newton-Raphson algorithm was used here after a few initial Deming-Stephan iterations to obtain good starting values. Both algorithms are implemented in the computer program LOGIMO (Kelderman & Steen, 1993).

A complicated problem for both algorithms is the computation of the expected sufficient statistics $E(R_n)$ and $E(S_n)$. Efficient solutions for the PCM were described by Andersen (1977) and Fischer (1974). For more general discrete data models, efficient computational procedures have been described by Kelderman (1992).

Testing Multidimensionality of Partial-Credit Models

For exponential family models based on discrete data (Equation 26), various goodness-of-fit statistics are available. Bishop et al. (1975) and Haberman (1979) provided a complete account of these statistics [e.g., Pearson's statistic ($X^2$) and the likelihood-ratio statistic ($G^2$)]. These statistics are asymptotically distributed as $\chi^2$'s with degrees of freedom (df) equal to the difference between the number of possible response patterns and the number of independent model parameters. If, however, the expected counts of the response patterns become too small, the approximation of the distribution of the overall goodness-of-fit statistic $X^2$ and $G^2$ by a $\chi^2$ distribution becomes poor (Koehler, 1977; Lancaster, 1961), although the distribution of $X^2$ is generally closer to $\chi^2$ than to $G^2$ (Cox & Placket, 1980; Larnz, 1978). The usual criterion for the size of the expected counts is 5; however, if the distribution is smooth the minimum expected count could be as small as 1 (Cochran, 1952, 1954).

If the minimum expected counts are too small, the model of interest can be tested against an alternative model that contains the PCM of interest as a special case, but that also contains parameters describing a particular deviation. This alternative model will be called a diagnostic model. For the size concept data discussed below, a possible diagnostic model for the 2PCM would be a 2PCM with different item response parameters in each age group.

A statistic that can be used to compare a PCM with such a diagnostic model is the likelihood-ratio statistic. Let $L = \log P(X|T)$ be the loglikelihood of the model of interest and $L^*$ be the loglikelihood of the diagnostic model. The likelihood-ratio test statistic is

$$LR = -2(L - L^*),$$

which is asymptotically distributed as $\chi^2$ with df equal to the difference in the number of linearly independent parameters of both models (Rao, 1973, pp. 418-420). Haberman (1977) showed that the likelihood-ratio statistic has good asymptotic properties.

**Example Applications**

**Example 1: Size Concept Data**

Data

Van Kuyk (1988) collected data from an observation program of 4-6.5 year-old children. The program tested skills prerequisite for arithmetic abilities. One subtest was reanalyzed here because it requires partial-credit scoring. The test measured the application of size concepts such as "long-short," "high-low," "thick-thin," and "wide-narrow." For example, in Item 6, a figure displaying four identical girls with successively shorter skirts was shown to the child. The test administrator pointed to the figure and said "Here you see some skirts. They gradually become a bit ..." and the examinee supplies the answer. Answers were rated incorrect if the answer was cognitively incorrect (e.g., "different" rather than "short"). Answers were rated correct if
the correct size concept was given (i.e., “long-short”) and was correctly applied [e.g., “short(er)" rather than “long(er)"]. Linguistic errors, such as “small" rather than “smaller," did not count as long as the answer was cognitively correct. Small children may be unable to produce the correct specific concept (e.g., “long-short"), but may use the general size concept “big-small" instead. If “big-small" was correctly applied [e.g., “the skirt is small(er)" the answer was rated partially correct.

The analysis of these data focused on the question of whether the ability to apply the “big-small" general concept was the same latent trait as the ability to apply the specific dimensional size concept “long-short.” It may be hypothesized that these traits are not identical because the first activity may be based solely on the perceptual saliency of a picture element (e.g., the skirt), whereas the second process depends on the correct identification of a particular figural property of that element (e.g., vertical length).

Each of the \( N = 263 \) persons responded to \( n = 15 \) items. The random response \( X_{ij} \) of fixed person \( j \) (\( j = 1, ..., N \)) on fixed item \( i \) (\( i = 1, ..., n \)) was modeled. In this example, this response was scored with values \( x_{ij} = 0 \) for an incorrect response, \( x_{ij} = 1 \) for a partially correct response, and \( x_{ij} = 2 \) for a correct response, where \( x = 1, ..., r \).

The sample was divided into three age groups: for ages 4–5 there were 66 children; ages 5–5.5, 132 children; and ages 5.5–6, 65 children. The likelihood equations of the PCM and MPCM estimated here were solved simultaneously in all three subpopulations, and the item response parameters \( \delta \) were set equal over subgroups. The \( \eta \) parameters were allowed to vary between subgroups. This required a slight extension of the likelihood equations. To accomplish this, a subgroup subscript—\( I \)—was added to \( S \) and \( \tau \) in the likelihood equations: \( S_{ijl} = E(S_{il}) \).

\section*{Results}

\textit{PCM.} Table 1 shows the item threshold parameter estimates \( (\delta_{ij}) \) of the IPCM for the size concept data. The parameters \( \delta_{ij} \) pertain to the log odds of the partially correct response (\( x = 1 \)) relative to the incorrect response (\( x = 0 \)), the parameters \( \delta_{ij} \) pertain to the log odds of the correct response (\( x = 2 \)) relative to the partially correct response (\( x = 1 \)), and the parameters \( \delta_{ij} + \delta_{ij} \) pertain to the log odds of the correct response (\( x = 2 \)) relative to the incorrect response (\( x = 0 \)). Note that the identifying constraint in Equation 2 fixes the sum of all parameters \( \delta_{ij} + \delta_{ij} \) to be equal to 0, so that the mean of the parameters \( \delta_{ij} + \delta_{ij} \) is 0.

Table 1 shows that the mean (−.37) of \( \delta_{ij} \) for the partially correct responses was lower than 0. This means that, on average, it was less difficult to give a “big-small" response than to give a correct specific size concept (e.g., “shorter" relative to an incorrect response. For example, in Item 6 it was relatively easy to see that the rightmost skirt was “smaller" (\( \delta_{61} \)) but it was more difficult to apply the correct size concept "shorter" both relative to the incorrect response (\( \delta_{61} + \delta_{62} = 1.36 \)) and to the partially correct response (\( \delta_{63} = 1.35 \)).

Item 8 is an item for which the reverse is true. For Item 8, a figure is shown with three ladders. On each of the ladders is an identical child standing on a rung, but the children are standing on lower positions on consecutive ladders. In Item 8 it is relatively difficult to give the “big-small" answer (\( \delta_{81} = 2.78 \)) rather than the incorrect answer, but easier to give the correct answer (\( \delta_{82} = -1.60 \) “lower" rather than the “big-small" answer. This seems rather obvious because in the pictures there are no elements differing in size, only position, whereas in Item 6 the skirts clearly differ in size.

To get the fully correct answer, the specific type of size concept also seemed to matter. Concepts describing the vertical dimension—“low, less, lower,” and “higher”—seemed to be more difficult than the size concepts that were not associated with a particular direction such as “long, short, thin, thickest, longest,” and “thinnest.”

\textit{MPCM.} A 2PCM was also specified for these data. It was hypothesized that individual differences in the ability to apply the general size concept “big-small" were qualitatively different from the individual differences in the ability to apply the specific size concept (e.g., “long-short”). Therefore, the MPCM was
Table 1
Parameter Estimates of the IPCM and the 2PCM for the Size Concept Data

<table>
<thead>
<tr>
<th>Item and Content</th>
<th>IPCM</th>
<th>2PCM</th>
<th>Normed as in IPCM</th>
<th>Normed to Zero Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>δ1</td>
<td>δ2</td>
<td>δ1 + δ2</td>
<td>δ1+δ2</td>
</tr>
<tr>
<td>(Long)</td>
<td>-0.79</td>
<td>-1.28</td>
<td>-2.07</td>
<td>-0.69 + -1.39 + -2.07</td>
</tr>
<tr>
<td>(Low)</td>
<td>-0.59</td>
<td>2.34</td>
<td>2.93</td>
<td>0.52 + 2.40 + 2.92</td>
</tr>
<tr>
<td>(Short)</td>
<td>-0.24</td>
<td>-1.17</td>
<td>-2.31</td>
<td>-0.24 + -1.19 + -2.31</td>
</tr>
<tr>
<td>(A lot)</td>
<td>-.30</td>
<td>-1.27</td>
<td>-1.97</td>
<td>-.41 + -1.39 + -1.97</td>
</tr>
<tr>
<td>(Thin)</td>
<td>-0.83</td>
<td>-0.86</td>
<td>-1.69</td>
<td>-0.83 + -0.95 + -1.70</td>
</tr>
<tr>
<td>(Shorter)</td>
<td>0.01</td>
<td>1.55</td>
<td>1.56</td>
<td>0.01 + 1.51 + 1.58</td>
</tr>
<tr>
<td>(Less)</td>
<td>.61</td>
<td>1.33</td>
<td>1.94</td>
<td>.58 + 1.36 + 1.94</td>
</tr>
<tr>
<td>(Lower)</td>
<td>2.78</td>
<td>-1.60</td>
<td>1.18</td>
<td>2.92 + -1.73 + 1.19</td>
</tr>
<tr>
<td>(Thickest)</td>
<td>.13</td>
<td>-0.21</td>
<td>-0.08</td>
<td>0.18 + -0.25 + -0.06</td>
</tr>
<tr>
<td>(Higher)</td>
<td>-0.85</td>
<td>2.17</td>
<td>1.32</td>
<td>-0.97 + 2.30 + 1.33</td>
</tr>
<tr>
<td>(Longest)</td>
<td>-0.84</td>
<td>.11</td>
<td>-.73</td>
<td>-0.82 + .11 + -.71</td>
</tr>
<tr>
<td>(Thickest)</td>
<td>-0.98</td>
<td>-.51</td>
<td>-1.49</td>
<td>-0.98 + -.56 + -1.49</td>
</tr>
<tr>
<td>(Most)</td>
<td>-1.35</td>
<td>1.98</td>
<td>.63</td>
<td>-1.47 + 2.12 + .65</td>
</tr>
<tr>
<td>(Thinnest)</td>
<td>-.03</td>
<td>-.89</td>
<td>-.92</td>
<td>.06 + -.98 + -.92</td>
</tr>
<tr>
<td>(Highest)</td>
<td>-2.30</td>
<td>3.11</td>
<td>.81</td>
<td>-2.45 + 3.29 + .84</td>
</tr>
<tr>
<td>Mean</td>
<td>-.37</td>
<td>.37</td>
<td>0.00*</td>
<td>-.37* + .37* + 0.00*</td>
</tr>
</tbody>
</table>

*Fixed.

specified such that q = 1, 2 and x = 0, 1, 2 with weights w_{q,x} = 1 if q = x and 0 otherwise.

The item parameter estimates for the 2PCM are also given in Table 1. In the 2PCM, the origin of both latent dimensions was indeterminate. To remove this indeterminacy, means of the item parameters for each latent dimension were set equal to a constant. In Table 1, the parameter estimates of the 2PCM are displayed for two choices of identifying constants (0.00, 0.00) and (-.37, .37) corresponding to the mean values found under the IPCM. The latter normalization simplifies comparison with the IPCM. Table 1 shows that the parameter estimates of the 2PCM were approximately the same as in the IPCM. For example, for Item 15 (δ_{1,1} and δ_{1,2}) were -2.30 and 3.11 for the IPCM and -2.45 and 3.29 for the 2PCM, respectively.

Goodness of fit. To select between the models, their goodness of fit was compared. The number of possible response patterns for the size concept data (315) was too large to use the overall X^2 or G^2 goodness-of-fit statistics. Therefore, the likelihood ratio statistic of Equation 27 was used to compare the fit of different models. Four models were considered: two versions of the IPCM and two versions of the 2PCM. The first version had threshold parameters that were the same across age groups and the second version had threshold parameters that differed across age groups. Table 2 gives -2 times the likelihood and the number of parameters of each of these models. These numbers were used to compute the likelihood ratio statistic of Equation 27. Comparing the fit of the IPCM with equal (invariant) parameters over subgroups with the IPCM with different (dependent) parameters over subgroups (Row 1), a large likelihood ratio statistic was obtained (3,436.51 - 3,280.01 = 156.50) relative to the df (57). Therefore, the former model was rejected against the latter at the .05 significance level.

Example 2: Progressive Matrices Items

Method

Data. Data from 1,061 persons to the Standard Raven Progressive Matrices Test (Raven, Raven, & Court, 1991) collected were reanalyzed (Vodegel Matzen, 1994). The Progressive Matrix Test is a nonverbal test that is usually assumed to measure the single trait of analytic intelligence.

The fit of several PCMs was compared and the best-fitting model was selected. Because chance capitalization may affect the results, the findings were cross-validated on an independent sample. To do this, the sample
Table 2

-2 Times the Likelihood (-2L) and Number of Parameters (No.)
for the IPCM and the 2PCM

<table>
<thead>
<tr>
<th>Model</th>
<th>Subgroup Invariant Parameters</th>
<th>Subgroup Dependent Parameters</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2L</td>
<td>No.</td>
<td>-2L</td>
</tr>
<tr>
<td>IPCM</td>
<td>3,436.51</td>
<td>108</td>
<td>3,280.01</td>
</tr>
<tr>
<td>2PCM</td>
<td>3,298.51</td>
<td>152</td>
<td>3,141.92</td>
</tr>
<tr>
<td>Difference</td>
<td>138.00</td>
<td>44</td>
<td>138.09</td>
</tr>
</tbody>
</table>

was randomly divided into an exploration sample \((N = 511)\) and a cross-validation sample \((N = 550)\).

Figure 2 shows a typical matrix item. To protect the security of the actual Raven items, an isomorph (which uses the same rules but different figural elements and attributes) is shown here. The items have eight answer alternatives. The person is asked to select the alternative that fits the matrix best. The answers can be found by scanning certain rules; some items require different rules than others. In the item in Figure 2, the height of the square diminishes in each row and the width of the square increases in each column. Both rules are called quantitative pairwise progression (Rule 1). Alternative 5 is the correct response—it has the largest width and the smallest height. Alternatives 1, 4, 7, and 8 are partially correct responses: either the height is correct and the width is incorrect (1 and 8), or the width is correct and the height is incorrect (4 and 7). The remaining responses (1, 3, and 6) are incorrect responses.

To illustrate the use of MPCMs in studying the dimensionality of matrix items, six items were selected \((n = 6)\): three items—C6, C7, and C10—were based on Rule 1 and three items—D7, D8, and D9—used a second rule (Rule 2). Table 3 shows for each response alternative whether it was incorrect (denoted by \(x = 0\)), partially correct \((x = 1)\), or fully correct \((x = 2)\).
Table 3
Scoring of Responses to Six Matrix Items Into the Categories Incorrect \( (x = 0) \), Partially Correct \( (x = 1) \), and Correct \( (x = 2) \)

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (C6)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 (C7)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3 (C10)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4 (D7)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5 (D8)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 (D9)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Models.** As multidimensional alternatives to the ipcm (Model A), several MPCMs may be specified for these data. One alternative hypothesis (Model B) is based on the idea that finding two instances of the rule rather than one requires more cognitive ability and not necessarily the same type as finding one instance only. Sternberg (1990, p. 121) discussed control processes that are involved in the execution of compounded cognitive tasks—such as a person keeping track of their place in task performance. It may be hypothesized that finding both instances of the rules invokes such a type of ability on which persons may differ. This hypothesis can be modeled by specifying a separate latent trait pertaining to the fully correct response (see Figure 1b).

Another alternative hypothesis (Model C) is that finding Rule 1 and Rule 2 instances pertain to different cognitive abilities. This hypothesis may be modeled by specifying a different latent trait for Items C6, C7, and C10 than for Items D7, D8, and D9.

Table 4 gives the B weight specifications for Rule 1 items and Rule 2 items of four MPCMs. Model B is two-dimensional within each item; that is, the fully correct response involves an additional latent trait corresponding to a specific metacognitive ability that is necessary to correctly combine two rule instances. Table 4 shows that in Model B both Rule 1 and Rule 2 items pertain to the same latent traits.

Model C is like Model A in that within each item only one latent trait is specified (see also Figure 1a). The difference between A and C is that different latent traits are specified for the items that involve Rule 1

### Table 4

<table>
<thead>
<tr>
<th>Model, Interpretation</th>
<th>Rule 1 Items</th>
<th>Rule 2 Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( x )</td>
<td>( x )</td>
</tr>
<tr>
<td>Model A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1: ) Find Rules General</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Model B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1: ) Find Rules General</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( \theta_2: ) Metacomponent</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Model C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1: ) Find Rules of Type 1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( \theta_2: ) Find Rules of Type 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Model D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta_1: ) Find Rules of Type 1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( \theta_2: ) Find Rules of Type 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \theta_1: ) Metacomponent</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
(Items 1, 2, and 3) than for the items that involve Rule 2 (Items 4, 5, and 6). Finally, Model D has both multidimensionality within and across the items. This model has three latent traits: one latent trait corresponding to Rule 1, one corresponding to Rule 2 (as in Model C), and one corresponding to the metacomponent (as in Model B).

Results

To test unidimensionality within the item responses, Model A was compared with Model B and Model C was compared with Model D using a likelihood ratio (LR) test. Both comparisons [\( LR(A, B) = 102, df = 13; LR(C, D) = 132, df = 8 \)] were significant at the .05 significance level. Thus, the null hypothesis of a PCM within each item was rejected in favor of the 2PCM.

To test the hypothesis that Rule 1 items pertained to a different latent trait than Rule 2 items, Model A was compared with Model C and Model B was compared with Model D. These comparisons were also highly significant [\( LR(A, C) = 147, df = 34; LR(B, D) = 177, df = 29 \)], indicating that both types of items pertained to different latent traits.

Table 5 shows item fit for Model D. In Table 5, Model D’s \( \chi^2 \) statistics are given for each combination of item (i) and \( \theta \) for which weights were specified in the model. Comparing these \( \chi^2 \) statistics with their \( df \), no significant outcomes were found that would lead to the rejection of Model D for the exploration sample. Therefore, it was concluded that Model D, specifying a different latent trait for each of the rules and a common latent trait for the metacognitive ability in combining the rules, fit the data. Table 5 also shows the cross-validation results of the \( \chi^2 \) statistics for Model D. These results also indicated a good fit to the data. In the cross-validation sample, none of the statistics exceeded the \( \alpha = .05 \) critical value of 11.1 (\( df = 5 \)) or 19.7 (\( df = 11 \)).

Table 6 shows the parameter estimates for Model D. For each latent trait, one parameter was fixed to 0 to make the model identifiable: \( \delta_1 \) fixed the scale of the first latent trait (Rule 1), \( \delta_2 \) fixed the scale of the second latent trait (Rule 2), and \( \delta_3 \) fixed the scale of the third latent trait (metacomponent). The 8 parameters for consecutive odds as well as the \( \phi \) parameters are given. Note that the \( \delta \)s can only be compared if they pertain to the same latent trait. For example, the parameters \( \delta_{i2} \) (\( i = 1, \ldots, n \)) all pertain to the third latent trait and can, therefore, be compared. Table 6 shows that Items 1 and 2 were easier on the third latent trait than Items 3–6 (their parameters \( \delta_{i3} = 0.00 \) and \( \delta_{i3} = .07 \) were smaller than \( \delta_{i3} = 1.25, \delta_{i3} = 1.02, \delta_{i3} = .79, \) and \( \delta_{i3} = 1.24 \)).

Discussion

By applying MPCMs to van Kuyk’s size concept data and the Raven Progressive Matrices Test, it was shown that MPCMs can be used to test hypotheses about the dimensionality of polytomous test data. The
Table 6
Parameter Estimates ($\phi$) of the Three-Dimensional PCM (Model D) for Progressive Matrices Data for the Total Sample

<table>
<thead>
<tr>
<th>Item, $\delta$, and $\phi$</th>
<th>Item Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (C6)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.00*</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.00*</td>
</tr>
<tr>
<td>2 (C7)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>-0.27</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.27</td>
</tr>
<tr>
<td>3 (C10)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.28</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-1.28</td>
</tr>
<tr>
<td>4 (D7)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.00*</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.00*</td>
</tr>
<tr>
<td>5 (D8)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-0.05</td>
</tr>
<tr>
<td>6 (D9)</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.62</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-0.62</td>
</tr>
</tbody>
</table>

*Fixed to 0.

2PCM of the size concept example was a special case of the general MPCM in which all partially correct scores pertained to one latent trait and all fully correct scores pertained to one latent trait. In that case the MPCM is in fact equivalent to Rasch’s original multidimensional model (Andersen, 1973b). To show this, let $r_i = s$ and $w_{qiy} = 1$ for $q = y$, and 0 otherwise; then:

$$\ln(P_{i\mu}/P_{i\nu}) = \sum_{y=1}^{Q} \sum_{q=1}^{Q} w_{qiy} \theta_{y} - \sum_{y=1}^{Q} \delta_{y} - \sum_{y=1}^{Q} \delta_{y}^* = 0^* - \delta_{y}^*$$

which is the MPCM. If, however, there are different specifications of $w_{qiy}$, such as in the Progressive Matrices example, this may no longer be the case.

To test the goodness of fit of the items in a certain MPCM, item by weight-sum goodness-of-fit statistics were used. Further research is needed on how to combine these statistics into one overall goodness-of-fit statistic. For the dichotomous Rasch model, van den Wollenberg (1979, 1982) proposed the $Q_i$ statistic, which is a weighted sum of the item by sum-score statistics $X^2_i$ for the dichotomous Rasch model. Although simulations have shown that the distribution of the statistic is close to $\chi^2$, no proof of this is available to date. Glas (1988) derived a similar statistic for which he proved that it is asymptotically a $\chi^2$. The statistic involves the inversion of a second derivative matrix that is more difficult to obtain. It is worthwhile to study a generalization of this statistic for MPCMs.

Appendix A

Kelderman (1992) described Deming-Stephan and Newton-Raphson algorithms to solve the likelihood equations iteratively. Let $K_{xi} = R_{xi} - R_{xi-1}$ be the number of persons with response $x$ on item $i$. Starting with parameter values of 0, the Deming-Stephan algorithm (also called iterative proportional fitting) yields the MLES by repeated application of...
and where "old" and "new" denote the parameter values before and after iteration, respectively. The Sex parameters can then be obtained from

\[ \text{Appendix B} \]

The Newton-Raphson algorithm is faster but more sensitive to the choice of starting values than the Deming-Stephan algorithm. It is based on the iteration

\[
\begin{bmatrix}
\hat{\phi}^{(\text{new})} \\
\hat{\phi}^{(\text{new})}
\end{bmatrix} =
\begin{bmatrix}
\phi^{(\text{old})} \\
\phi^{(\text{old})}
\end{bmatrix} +
\begin{bmatrix}
K - E(K)^{(\text{old})} \\
S - E(S)^{(\text{old})}
\end{bmatrix}
\begin{bmatrix}
H^{(\text{old})}
\end{bmatrix}^{-1},
\]

where \( \phi = \phi^{(\text{old})} \), \( x = (x) \), \( K = (K_x) \), \( S = (S_x) \), and \( H \) is the Hessian matrix. Let \( M_{i1} = (M_{ij}) \), where \( M_{ij} \) is the number of persons with response \( x \) on item \( i \) and response \( x' \) on item \( i' \). Define \( M_{i2} = M_{2i} = (M_{ix}) \) similarly. Furthermore, let \( M_{22} = \text{diag}(S) \). The Hessian is then

\[
H = E\left(\begin{bmatrix}
M_{11} \\
M_{12} \\
M_{21} \\
M_{22}
\end{bmatrix}\right).
\]

References


Haberman, S. J. (1977). Log-linear models and frequency tables with small expected cell counts. Annals of Sta-
statistics, 5, 1148–1169.

**Author’s Address**

Send requests for reprints or further information to Henk Kelderman, Division of Work and Organizational Psychology, Faculty of Psychology and Pedagogics, Vrije Universiteit Amsterdam, van de Boechorststraat 1, 1018 BT Amsterdam, The Netherlands. Email: h.kelderman@psy.vu.nl.