PSYCHOMETRIKA—VOL. 59, NO. 2, 177–192
JUNE 1994

AN EXTENSION OF THE PARTIAL CREDIT MODEL WITH AN APPLICATION TO THE MEASUREMENT OF CHANGE

GERHARD H. FISCHER AND IVO PONOCNY

UNIVERSITY OF VIENNA

The partial credit model is considered under the assumption of a certain linear decomposition of the item × category parameters δ_{ih} into "basic parameters" α_j. This model is referred to as the "linear partial credit model". A conditional maximum likelihood algorithm for estimation of the α_j is presented, based on (a) recurrences for the combinatorial functions involved, and (b) using a "quasi-Newton" approach, the so-called Broyden-Fletcher-Goldfarb-Shanno (BFGS) method; (a) guarantees numerically stable results, (b) avoids the direct computation of the Hesse matrix, yet produces a sequence of certain positive definite matrices $\mathbf{B}_k$, $k = 1, 2, \ldots$, converging to the asymptotic variance-covariance matrix of the $\hat{\alpha}_j$. The practicality of these numerical methods is demonstrated both by means of simulations and of an empirical application to the measurement of treatment effects in patients with psychosomatic disorders.

Key words: Partial credit model, Rasch model, item response theory, measurement of change, assessment of treatment effects.

1. Motivation

In their daily work with psychology students writing a thesis, the authors of the present paper continuously experience a lack of appropriate methods for assessing effects of treatments based on polytomous ordered responses to ability, personality, attitude, or self-rating items, or to ratings of clinical symptoms. Most statistical methods for qualitative data do not seem quite adequate because they make no allowance for individual differences in terms of person parameters. For this reason, we propose an IRT approach to the measurement of change. Such a methodology for *dichotomous* items has been developed and applied successfully in the past 15 years (see Fischer, 1989, 1991, and earlier references therein), but similar methods for *polytomous* items are largely missing.

One recent paper (Fischer & Parzer, 1991a) extended the "rating scale model" (RSM) of Andrich (1978a, 1978b)—a special case within the family of polytomous Rasch models (Rasch, 1961, 1965, 1967)—by imposing a linear structure on the item parameters. This "linear rating scale model (LRSM)" was applied to clinical data on the effects of psychotherapy given to patients with a "functional syndrome". This study, however, showed that the LRSM, which characterizes each response probability in terms of one item parameter β_i, one category parameter ω_h, and one person parameter θ_v, is too restrictive for obtaining a good fit to data of that complexity. The present paper therefore is based on the "partial credit model" (PCM; Andersen, 1983; Masters, 1982), which assigns one independent parameter β_{ih} to each item I_i × category C_h combination and therefore is much more flexible. Again a linear structure of the parameters is introduced, in this case embedded in the PCM. This extension of the

The authors thank one anonymous reviewer for his constructive comments. Moreover, they thankfully acknowledge financial support by the Österreichische Nationalbank (Austrian National Bank) under Grant No. 3720.

Requests for reprints or for the LPCM software (please send a formatted 3 1/2 inch diskette) should be directed to Gerhard H. Fischer, Institut für Psychologie, Liebiggasse 5, A-1010 Wien (Vienna), Austria, e-mail A6212DAC@AWIUNI11.BITNET.

0033-3123/94/0600-92037$00.75/0

PCM might on first sight appear to be a minor change compared to the LRSM because some formulas look similar, but any closer inspection will reveal substantial differences. Firstly, the LPCM is a more general structure allowing one to formulate and test empirically a host of new hypotheses, that is, the LRSM is just a special case of the LPCM. Second, the LPCM has many more parameters; their estimation poses nontrivial problems and requires a new algorithm for the computation of a conditional maximum likelihood (CML) solution. This can be illustrated by means of the following example: Consider, for instance, a self-rating scale administered at only one time point, consisting of $k = 30$ items with $m + 1 = 5$ ordered response categories ($C_0, C_1, \ldots, C_m$; $m = 4$). The LPCM comprises $km - 1$ item $\times$ category parameters, i.e., 119 parameters have to be estimated under the CML approach (see sec. 2). In the LRSM, the linear part of the exponential function is $h(\theta_v + \beta_i) + \omega_h$, so that $k - 1 = 29$ item parameters plus $m - 1 = 3$ category parameters, that is, 32 parameters have to be estimated. In other words, in this example the PCM would have app. 3.7 times as many parameters as the RSM. Since the computing time for one iteration in the Newton-Raphson algorithm employed by Fischer and Parzer (1991a, 1991b), depending mainly on the second-order partial derivatives and on a matrix inversion, increases approximately by the third power of the number of parameters, the corresponding numerical computations in the PCM would run about $3.7^3 \approx 51$ times as long. This expense seems to have deterred authors from using the CML method in the PCM. Therefore, we now develop a more efficient CML algorithm for the LPCM.

2. The LPCM

The PCM is defined by the equation

$$P(A_{vih} = 1 | S_v, I_i) = \frac{\exp(h\theta_v + \beta_{ih})}{\sum_{t=0}^{m_i} \exp(t\theta_v + \beta_{it})}, \tag{1}$$

for $i = 1, \ldots, k$ items I_i and $h = 0, \ldots, m_i$ response categories C_h, where A_{vih} is the polytomous response variable of subject S_v on item I_i, with realizations $a_{vih} = 1$ if S_v chooses category C_h, and $a_{vih} = 0$ otherwise; all A_{vih} are assumed to be "locally" independent;

$m_i + 1$ is the number of response categories of item I_i, $C_0, \ldots, C_m$;
β_{ih} is the attractivity (or easiness) of category C_h of item I_i;
θ_v is the position of subject S_v on the underlying unidimensional latent trait.

For making the model identifiable, the parameters are normalized as follows:

$$\beta_{i0} = 0 \quad \text{for} \quad i = 1, \ldots, k, \qquad \sum_{i=1}^{k} \sum_{t=1}^{m_i} \beta_{it} = 0. \tag{2}$$

Therefore, the number of independent parameters in the PCM is $\sum_i m_i - 1$, or, if all $m_i = m$, $km - 1$. As is wellknown, the sufficient statistics for the θ_v are $r_v = \sum_i \sum_{t=0}^{m_i} t a_{vit}$; and the sufficient statistics for the item parameters are the marginal vectors $\mathbf{s}_i = (s_{i0}, \ldots, s_{im_i})$ with elements $s_{ih} = \sum_v a_{vih}$ (see Masters, 1982).

In the present paper, the PCM is considered under the linear restrictions

$$\beta_{ih} = \sum_{j=1}^{p} w_{ihj}\alpha_j + hc, \tag{3}$$

for $i = 1, \ldots, k$ and $h = 1, \ldots, m_i$, where

c is a normalization constant,
α_j, $j = 1, \ldots, p$, are "basic parameters" measuring the effects of structural item properties or of experimental conditions on the responses, and
w_{ihj} are given weights of the α_j.

The PCM (1) combined with (3) will be called the "linear partial credit model" (LPCM). The basic parameters α_j allow various interpretations depending on the particular application: They may represent effects of certain cognitive operations involved, or the effect of learning or treatment on a trait, or the effect of communications or advertising on an attitude measured by Likert-style items, etcetera. Our main concern is the application of the LPCM to repeated measurement designs for assessing the effects of treatments given to some or all subjects.

The LPCM has been considered by Glas and Verhelst (1989) within the marginal maximum likelihood (MML) framework of parameter estimation. In the present paper, we shall deal with the LPCM only on the basis of the CML approach which, as we shall show, yields an excellent framework for various conditional likelihood ratio tests of composite hypotheses regarding the treatment effects.

In order to cope with any repeated measurement design, we introduce the proven notion of "virtual items" as opposed to "real items": a real item is a certain task or question, characterized in (3) by item parameters β_{ih}, $h = 0, \ldots, m_i$, as usual; a virtual item (V-item), on the other hand, is a technical concept defined only within the framework of IRT models, namely, a situation characterized by the presentation of a real item under specified conditions. These conditions may change the response behavior of the subjects. For formal convenience, we shall consider the person parameters θ_v fixed constants and describe change in terms of the item parameters alone. This implies that the parameters of an item may change depending on the condition or time point at which the response is given. One real item therefore generates as many V-items as there are combinations of time points $\times$ conditions (or treatments) in the actual design. For distinguishing real items I_i from V-items, the latter will be denoted I_l^*, $l = 1, \ldots, u$. A V-item I_l^* is formally characterized by V-item parameters denoted δ_{lh}, $h = 0, \ldots, m_l$. If, as is mostly the case in treatment studies, the dosage of the treatments varies between subjects (or subject groups), subsets of V-item parameters depend on each subject (or subject group).

Suppose, for instance, 15 real items, $I_1, \ldots, I_{15}$, are presented at three time points, T_1, T_2, T_3, to subjects of two treatment groups, the treatments being given in the intervals between testing occasions. This design generates $3 \times 15 = 45$ V-items for group G_1, $I_1^*, \ldots, I_{45}^*$, 15 of which, however, are identical with $I_1, \ldots, I_{15}$ (namely, $I_1^*, \ldots, I_{15}^*$ pertaining to time point T_1). Similarly, there are also 45 V-items for group G_2, the first 15 of which again are identical with real items $I_1, \ldots, I_{15}$. Hence, there are a total of $15 + 30 + 30 = 75$ V-items, which may be denoted I_l^*, $l = 1, \ldots, 75$. Table 1 illustrates the situation. This notation is most convenient for the derivations in section 3. For purposes of concrete applications, however, it often is more practical to replace the abstract index l by an index combination referring to one subject (or treatment group) $\times$ real item $\times$ time point. We shall make use of that possibility in sections 5 and 6. Using the V-items concept, any repeated measurement design may be treated as an *incomplete* design of an LPCM with V-item parameters δ_{lh}.

Table 1.

A structurally incomplete design with 15 real items and 75 V-items presented to subjects of two treatment groups, G_1 and G_2, at three time points, T_1, T_2, T_3. V-items $I_1^*, \ldots, I_{15}^*$ are identical with real items $I_1, \ldots, I_{15}$. Therefore, G_1 and G_2 respond to the same V-items at T_1, but to different V-item sets at T_2 and T_3.

	V-items presented at		
	T_1	T_2	T_3
G_1	$I_1^*, \ldots, I_{15}^*$	$I_{16}^*, \ldots, I_{30}^*$	$I_{31}^*, \ldots, I_{45}^*$
G_2	$I_1^*, \ldots, I_{15}^*$	$I_{46}^*, \ldots, I_{60}^*$	$I_{61}^*, \ldots, I_{75}^*$

We now reformulate the LPCM in terms of V-item parameters δ_{lh},

$$\delta_{lh} = \sum_{j=1}^{p} w_{lhj}\alpha_j + hc, \tag{4}$$

for $l = 1, \ldots, u$ V-items I_l^* and $h = 1, \ldots, m_l$ categories, with otherwise the same notation as in (3). The parameters δ_{lh} are normalized like the β_{ih} in (2). Notice that (4) defines the δ_{lh} only for $h = 1, \ldots, m_l$, whereas $\delta_{l0} = 0$ for all l by normalization.

Suppose we have a design with a treatment and a control group, G_1 and G_2, and two time points, T_1 and T_2, respectively. If we want to hypothesize that the V-item parameters δ_{lh} are composed of (a) item parameters β_{ih} attached to the real items, $I_1, \ldots, I_k$, and (b) effects of treatments B_j given between T_1 and T_2, changing the latent trait parameters θ_v of the subjects of G_1, we set

$$\delta_{lh} = \beta_{ih} + hc \quad \text{for all } S_v \text{ at } T_1, \text{ with } l = i,\ i = 1, \ldots, k,$$

$$\delta_{lh} = \beta_{ih} + h\sum_{j=1}^{b} q_j\eta_j + h\tau + hc \quad \text{for } S_v \in G_1 \text{ at } T_2,\ l = i + k,\ i = 1, \ldots, k, \tag{5}$$

$$\delta_{lh} = \beta_{ih} + h\tau + hc \quad \text{for } S_v \in G_2 \text{ at } T_2,\ l = i + 2k,\ i = 1, \ldots, k.$$

Therein,

β_{ih} are the item parameters of the real items, $I_1, \ldots, I_k$,
q_j is the dosage of treatment B_j given between T_1 and T_2 to subjects of G_1,
η_j is the effect of one unit of treatment B_j (i.e., the amount of change induced in θ_v by one unit of the treatment),
τ is a "trend" parameter (i.e., the effect of a "treatment" given to all subjects, e.g., the effect of being tested with the same test twice), and
c is the usual normalization constant.

In this model, the total number of V-items is $3k$. Notice that (5) is just a special case of (4) where change can be absorbed entirely in the unidimensional θ_v parameters, whereas (4) makes allowance even for a different impact of the α_j on each category × item parameter δ_{lh}.

Two other special cases of (4) are the PCM and the RSM: For $m_l = m$, we get the PCM if we set $\delta_{lh} = \alpha_j$ with $j = (l - 1)m + h$, for $l = 1, \ldots, k$, in (2); and we get the RSM by setting $\delta_{lh} = h\beta_l + \omega_h$ for $l = 1, \ldots, k$ and $h = 0, \ldots, m$, so that the β_i and ω_h are the α_j in (4).

The linear equations (4) therefore constitute a very general structure sufficing for a host of applications. Thus we arrive at the following definition of the LPCM:

$$P[A_{vlh} = 1 | \theta_v, (\delta_{lh})] = \frac{\exp(h\theta_v + \delta_{lh})}{\sum_{t=0}^{m_l} \exp(t\theta_v + \delta_{lt})}, \tag{6}$$

subject to restrictions (4), for all V-items I_l^*.

3. CML Estimation for the LPCM

A particularly attractive method of parameter estimation for the PCM is the conditional maximum likelihood (CML) method: The likelihood of the responses of a subject, given his/her raw score r_v, does not involve the person parameters θ_v; the sum of the corresponding log-likelihoods can therefore be maximized as a function of the item parameters β_{ih} alone (Andersen, 1983, pp. 122–125; Glas, 1989, pp. 113–116; Wright & Masters, 1982, pp. 85–86). A theoretical and practical advantage of the CML approach over other methods is that likelihood ratio tests of a multitude of hypotheses can easily be carried out immediately after the estimation without further computations. (For examples, see section 6 below.)

For the present derivations care has to be taken that all operations are valid for *incomplete* data since structurally incomplete designs with V-items are considered. To that end, we extend the definition of the response variables A_{vlh} by setting $A_{vlh} = 0$, $h = 0, \ldots, m_l$, for all S_v who do *not* respond to I_l^* (or to whom I_l^* is not given). The role of this definition will be to delete all those terms from the likelihood function corresponding to combinations of subjects × V-items where no observations are made.

For simplicity, let subjects be grouped by their treatment dosages, yielding "treatment groups G_g". (It suffices to consider cases of grouped data, since any case of individual dosages can be subsumed as having group size $n_g = 1$.) Now it follows immediately from the well-known form of the conditional likelihood in the PCM (see Andersen, 1983, p. 123; Masters, 1982) that

$$\ln L = \sum_g \left(\sum_{l=1}^{u} \sum_{h=1}^{m_l} s_{glh} \delta_{lh} - \sum_r n_{gr} \ln \gamma_{gr} \right), \tag{7}$$

where

$s_{glh} = \sum_{S_v \in G_g} a_{vlh}$ is the number of subjects in group G_g who have chosen category C_h of V-item I_l^*,
n_{gr} is the number of subjects in treatment group G_g who have score r, and
γ_{gr} is a combinatorial function of parameters $\varepsilon_{lh} = \exp(\delta_{lh})$ for $l = 1, \ldots, u$ and $h = 0, \ldots, m_l$, namely,

$$\gamma_{gr} = \sum_{\mathbf{x}|r} \prod_{l=1}^{u} \prod_{h=0}^{m_l} \varepsilon_{lh}^{x_{lh}}, \tag{8}$$

the summation in (8) running over all response patterns $\mathbf{x}$ compatible with raw score r now given by

$$r = \sum_{l=1}^{u} \sum_{h=1}^{m_l} h x_{lh}, \tag{9}$$

where the x_{lh} are defined analogously to the elements of the observed response patterns, a_{vlh}.

Consider the case that some V-item, say I_k^*, is *not* presented to subjects $S_v \in G_g$: then, by definition, $a_{vkh} = 0$ for $h = 0, \ldots, m_k$, and also $x_{kh} = 0$ for $h = 0, \ldots, m_k$ in all patterns $\mathbf{x}$. This implies that all terms $\varepsilon_{kh}^{x_{kh}}$, $h = 0, \ldots, m_k$, in (8) become 1, and furthermore $s_{gkh} = 0$ for $h = 0, \ldots, m_k$, in (7), such that omitted responses do not influence equations (7). Hence, (7) applies both to complete and incomplete data.

It is easy to show that the conditional likelihood of the LPCM—like any other polytomous Rasch model—defines an exponential family (Andersen, 1983). If an exponential family is in "minimal" (or "canonical") form, the log-likelihood function is convex and has at most one maximum in the interior of its domain. Whether that is the case for a concrete LPCM and a given data set, however, is not easy to verify because the answer depends jointly on the design (i.e., on the weights w_{lhj}) and on certain structural properties of the data. Jacobsen (1989) has given a general necessary and sufficient uniqueness condition for a larger class of exponential models comprising the LPCM, but—unfortunately—his condition is difficult to use because it involves a very large system of linear inequalities. Nevertheless, the general results on exponential families imply that, if the parameters have been properly normalized and the sample size is not extremely small, we can practically be assured that the conditional log-likelihood function will have a unique finite maximum.

For maximizing (7) the partial derivatives of (7) with respect to the basic parameters α_j, $j = 1, \ldots, p$, are set to zero,

$$\frac{\partial \ln L}{\partial \alpha_j} = \sum_{g} \sum_{l=1}^{u} \sum_{h=0}^{m_l} w_{lhj} \left(s_{glh} - \sum_{r} n_{gr} \gamma_{gr}^{-1} \frac{\partial \gamma_{gr}}{\partial \varepsilon_{lh}} \varepsilon_{lh} \right) = 0, \tag{10}$$

yielding estimates $\hat{\alpha}_j$ of the basic parameters subject to restrictions (4). (Note that, if item I_l^* is not given to group G_g, both s_{glh} and $\partial \gamma_{gr} / \partial \varepsilon_{lh}$ are zero for all h.)

For solving (10), the combinatorial functions γ_{gr} and $\partial \gamma_{gr} / \partial \varepsilon_{lh}$ have to be computed. Doing this by direct enumeration of the respective patterns $\mathbf{x}$, see definition (8), would not be practicable. Fortunately, there exist convenient recurrence formulae: Consider those V-items I_l^* to which subjects of group G_g give a response; let there be s of them, and let the set of these items be denoted $J = \{J_1, \ldots, J_s\}$ for convenience. For the combinatorial expression γ_{gr}, all products of parameters ε_{lh} are required which correspond to elements $x_{lh} = 1$ of patterns $\mathbf{x}$ of responses (given to the items in J) compatible with raw score r. Attaining raw score r on items $J_1, \ldots, J_s$ is possible by attaining any of the raw scores $r - h$, $0 \leq h \leq \min(r, m_s)$, on items $J_1, \ldots, J_{s-1}$, combined with raw score h on item J_s. Hence,

$$\gamma_{gr}(J_1, \ldots, J_s) = \sum_{h=0}^{\min(r,m_s)} \gamma_{g,r-h}(J_1, \ldots, J_{s-1})\varepsilon_{sh}, \tag{11}$$

which expresses $\gamma_{gr}(J_1, \ldots, J_s)$ in terms of γ-functions of the partial test $J_1, \ldots, J_{s-1}$ and of the parameters ε_{sh} of item J_s.

Setting $\gamma_{gr} = 1$ for $r = 0$, and $\gamma_{gr}(J_1) = \varepsilon_{1r}$ for $r = 0, \ldots, m_1$, (11) can be used as a recurrence for building up all γ-functions: The computation begins with $\gamma_{gr}(J_1, J_2) = \sum_{h=0}^{\min(r,m_2)} \gamma_{g,r-h}(J_1)\varepsilon_{2h}$ for $r = 0, \ldots, m_1 + m_2$. In each successive step, the $\gamma_{gr}(J_1, \ldots, J_t)$ are computed for all raw scores r that are obtainable on the first t items of the set J, $t = 1, 2, \ldots, s$. Fortunately, these computations are numerically very stable because only products of positive numbers are added.

A recurrence for obtaining the partial derivatives of the γ_{gr} immediately results from differentiating (11), which yields

$$\frac{\partial \gamma_{gr}(J_1, \ldots, J_s)}{\partial \varepsilon_{tk}} = \sum_{h=0}^{\min(r,m_s)} \left(\frac{\partial \gamma_{g,r-h}(J_1, \ldots, J_{s-1})}{\partial \varepsilon_{tk}} \varepsilon_{sh} + \gamma_{g,r-h}(J_1, \ldots, J_{s-1})\Delta_{st}\Delta_{hk} \right), \tag{12}$$

where Δ_{xy} denotes the Kronecker-Delta ($\Delta_{xy} = 1$ if $x = y$, and $\Delta_{xy} = 0$ otherwise).

The recurrent method for calculating the γ-functions and their derivatives via (11) and (12) is known as the "summation algorithm" because it only involves the summation of positive terms. It is the numerically most stable algorithm known for these combinatorial functions, but its drawback is that it is slow. A considerably faster method is the "difference algorithm" that was used already by Fischer and Allerup (1968; see also Fischer, 1974, pp. 241–244) for the dichotomous Rasch model. Since it was seen to become numerically unstable when tests are long (>40 items, say), it was later abandoned (for a discussion, see Gustafsson, 1980). With a minor modification, however, it can be used quite successfully, particularly so when suitably combined with the summation method, and it readily generalizes to polytomous items. This combination of both methods, yielding a fast and numerically accurate algorithm, works as follows:

The γ_{gr} are computed by means of (11) for all s V-items presented to group G_g. The partial derivative $\partial \gamma_{gr}/\partial \varepsilon_{tk}$ is

$$\frac{\partial \gamma_{gr}}{\partial \varepsilon_{tk}} = \gamma_{g,r-k}(J_1, \ldots, J_{t-1}, J_{t+1}, \ldots, J_s) =: \gamma^{(t)}_{g,r-k}.$$

Since

$$\gamma_{g1} = \gamma^{(t)}_{g0}\varepsilon_{t1} + \gamma^{(t)}_{g1}, \quad \gamma_{g2} = \gamma^{(t)}_{g0}\varepsilon_{t2} + \gamma^{(t)}_{g1}\varepsilon_{t1} + \gamma^{(t)}_{g2}, \text{ etcetera,}$$

it follows that

$$\gamma^{(t)}_{gr} = \gamma_{gr} - \gamma^{(t)}_{g,r-1}\varepsilon_{t1} - \cdots - \gamma^{(t)}_{g,r-q}\varepsilon_{tq}, \tag{13}$$

where $q = \min(r, m_t)$. Starting with $r = 1$, (13) is a complete recurrence relation for $r = 1, \ldots, r_{\max}$.

Clearly, since (13) involves subtractions, it *may* be affected by numerical error, but if the recurrence is applied for each item I_t separately, any numerical error occurring

for one index value t cannot spread to other values of t. In order to safeguard against numerical error, though, a check of the accuracy is recommended, to be carried out for each t separately. For that purpose, basically the same recurrence can be run top-down also, starting with $r = r_{\max}$ and reducing r by 1 in each step: From (13),

$$\gamma_{g,r_{\max}-l} = \gamma^{(t)}_{g,r_{\max}-l-m_t}\varepsilon_{tm_t} + \gamma^{(t)}_{g,r_{\max}-l-m_t+1}\varepsilon_{t,m_t-1} + \cdots + \gamma^{(t)}_{g,r_{\max}-l},$$

which yields the recurrence

$$\gamma^{(t)}_{g,r_{\max}-l-m_t} = \frac{1}{\varepsilon_{tm_t}}\left(\gamma_{g,r_{\max}-l} - \gamma^{(t)}_{g,r_{\max}-l} - \cdots - \gamma^{(t)}_{g,r_{\max}-m_t-(l-1)}\varepsilon_{t,m_t-1}\right). \quad (14)$$

Therefore, the top-down recurrence (14) can be carried out for $l = 0, \ldots, r_{\max} - m_t$, setting $\gamma^{(t)}_{gr} = 0$ for $r > r_{\max} - m_t$.

Thus, in principle, two values can be computed for each $\gamma^{(t)}_{gr}$, which then can be compared. Practically, however, it suffices to apply the bottom-up computation by means of (13) beginning with $r = 1$, until

$$\gamma^{(t)}_{gr} < \gamma^{(t)}_{g,r-m_t}\varepsilon_{tm_t},$$

(at $r = \bar{r}$, say), and then to continue with recurrence (14) from $r = r_{\max}$ to $r = \bar{r}$. Finally, these two "trial values" for $\gamma^{(t)}_{g\bar{r}}$ are compared. If they are close enough, all $\gamma^{(t)}_{gr}$ for this value of t are accepted; if not, all $\gamma^{(t)}_{gr}$ for this t are rejected and recomputed by the more stable summation algorithm using (11). (Our condition for accepting trial values was that the relative error of $\gamma^{(t)}_{g\bar{r}}$ be less than 0.00001.) Studies with simulated data showed that this combination of the difference and summation algorithms indeed accelerates the estimation procedure considerably, as compared to summation alone, without perceptible loss of accuracy.

A standard procedure for solving sets of nonlinear equations like (10) is the Newton-Raphson method. Because of the large number of parameters to be estimated in the present case (and the resulting size of the Hesse matrix), a quasi-Newton method based on successive approximations of the Hesse matrix is more efficient. This avoids the computation of the second-order partial derivatives of the γ-functions, but still yields an estimate of the asymptotic variance-covariance-matrix of the CML parameter solution. The method will be sketched in section 4.

For testing composite hypotheses on the basic parameters α_j, conditional likelihood ratio tests suggest themselves as a convenient technique ensuing the estimation. The model is estimated both under the H_0 and H_1 hypotheses, the conditional likelihood functions L_0 and L_1 are obtained from the estimates using (7), and the asymptotic χ^2 statistic $-2 \ln \lambda = 2(\ln L_0 - \ln L_1)$, with df equal to the difference of the respective numbers of independent (basic) parameters, is applied.

4. A Quasi-Newton Method

For avoiding the computation of the second-order partial derivatives, the so-called Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative approximation to the Hesse matrix (see Churchhouse, 1981, pp. 493–496; Kosmol, 1989, pp. 179–191) is used. The BFGS method can be considered as an approximation to the Newton-Raphson method. The latter is defined by the iterations

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k - \lambda_k \mathbf{H}_k^{-1} \nabla f_k, \quad (15)$$

where

$\boldsymbol{\alpha}_k$ is the k-th iterated column vector (here: of estimates of the basic parameters),
λ_k is a scalar determining the length of the k-th step,
$\mathbf{H}_k$ is the Hesse matrix taken at the point $\boldsymbol{\alpha}_k$, and
∇f_k denotes the gradient of the function f (here: of $-\ln L$) at $\boldsymbol{\alpha}_k$.

The BFGS method is presently considered the most successful quasi-Newton method for optimization problems. It replaces $\mathbf{H}_k^{-1}$ by a matrix $\mathbf{B}_k$ which is updated after each iteration according to the rule

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{v}_k\mathbf{s}_k' + \mathbf{s}_k\mathbf{v}_k'}{\langle \mathbf{y}_k, \mathbf{s}_k\rangle} - \frac{\langle \mathbf{v}_k, \mathbf{y}_k\rangle \mathbf{s}_k\mathbf{s}_k'}{\langle \mathbf{y}_k, \mathbf{s}_k\rangle^2}, \tag{16}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner vector product, $\mathbf{y}_k = \nabla f_{k+1} - \nabla f_k$, $\mathbf{s}_k = \boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k$, and $\mathbf{v}_k = \mathbf{s}_k - \mathbf{B}_k\mathbf{y}_k$.

The algorithm begins at an arbitrary starting point (e.g., $\boldsymbol{\alpha}_0 = \mathbf{0}$) with an arbitrary positive definite symmetric matrix $\mathbf{B}_0$ (e.g., the identity matrix), and iterates until the gradient is sufficiently near to zero in all components. If, at some step, the decrease achieved is too small, a gradient step is performed prior to the next iteration. In combination with a suitable step-length rule (we use the so-called "Armijo rule with extended step length", Armijo, 1966; Kosmol, 1989, p. 89), the BFGS method guarantees that all $\mathbf{B}_k$ are positive definite and that $-\mathbf{B}_k\nabla f_k$ is a direction of descent. This rule ensures that the descent cannot become too small at the beginning of the iteration or in the neighborhood of the solution; the step length is variable and is automatically controlled by the procedure itself. If the uniqueness conditions for the LPCM are satisfied, the $\boldsymbol{\alpha}_k$ converge to a minimum solution $\boldsymbol{\alpha}^*$ of $-\ln L$ at least in "Q-superlinear" form, that is, there exists a sequence of nonnegative integers c_k converging to zero such that, for all k, $\|\boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}^*\| \leq c_k\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}^*\|$. At the same time, the sequence of matrices $\mathbf{B}_k$ converges to the inverse of the Hesse matrix at $\boldsymbol{\alpha}^*$, so that the algorithm yields the same asymptotic covariance matrix one would get by means of the Newton-Raphson method.

5. Simulation Studies

A. *Large Sample Simulation*

In order to test the procedures outlined above, a set of data was generated for an LPCM with 20 items and 5 response categories per item, that is, the model comprised $4 \times 20 - 1 = 79$ item parameters β_{ih} uniformly distributed in the interval $(-1.5, +1.5)$. The subject parameters were randomly sampled from a standard normal distribution. The longitudinal design underlying the simulation comprised two time points, T_1 and T_2, and three groups of 2000 simulated subjects each (see Table 2). It was presumed that the first 10 items, $I_1, \ldots, I_{10}$, were given to all subjects at time point T_1, and the other 10, $I_{11}, \ldots, I_{20}$, at T_2. Such designs arise, for instance, in attitudinal or educational research when testing the subjects with the same items repeatedly appears to be inappropriate.

Within the framework of the LPCM, the model underlying the simulation can be formalized most conveniently by introducing the following notation: The V-item parameters δ_{lh} will be denoted δ_{vith}, that is, the abstract index l is now replaced by an index triplet (v, i, t) describing the subject (or group) $\times$ (real) item $\times$ time point combination which defines the conditions under which a response is given. Using this notation, the δ_{vith} are subject to the following linear decomposition:

Table 2.

Design with 20 real items, $I_1, \ldots, I_{20}$, two time points, T_1 and T_2, three groups, G_1, G_2, and G_3, and 40 V-items, underlying the simulation study. V-items $I_1^*, \ldots, I_{20}^*$ are identical with real items $I_1, \ldots, I_{20}$. All three groups respond to the same V-items at T_1, but to different V-item sets at T_2.

V-items presented at

	T_1	T_2
G_1	$I_1^*, \ldots, I_{10}^*$	$I_{11}^*, \ldots, I_{20}^*$
G_2	$I_1^*, \ldots, I_{10}^*$	$I_{21}^*, \ldots, I_{30}^*$
G_3	$I_1^*, \ldots, I_{10}^*$	$I_{31}^*, \ldots, I_{40}^*$

$$\delta_{vi1h} = \beta_{ih} \quad \text{for all } S_v \text{ at } T_1; \text{ real items } I_i,\ i = 1, \ldots, 10; \tag{17}$$

$$\delta_{vi2h} = \beta_{ih} \quad \text{for } S_v \in G_1 \text{ at } T_2; \text{ real items } I_i,\ i = 11, \ldots, 20; \tag{18}$$

$$\delta_{vi2h} = \beta_{ih} + h\eta_1 \quad \text{for } S_v \in G_2 \text{ at } T_2; \text{ real items } I_i,\ i = 11, \ldots, 20, \text{ i.e., V-items } I_l^* \text{ for } l = 21, \ldots, 30; \tag{19}$$

$$\delta_{vi2h} = \beta_{ih} + h\eta_2 \quad \text{for } S_v \in G_3 \text{ at } T_2; \text{ real items } I_i,\ i = 11, \ldots, 20, \text{ i.e., V-items } I_l^* \text{ for } l = 31, \ldots, 40. \tag{20}$$

The treatment effects were set to $\eta_1 = 1.0$ for group G_2, and $\eta_2 = 2.0$ for group G_3. Group G_1 was considered the Control Group with no treatment (= no change). (If, however, some "global" change or "trend effect" independent of the treatments had occurred, it would have been absorbed into the β_{ih} parameters of items $I_{11}, \ldots, I_{20}$, such that any global change would have been confounded with some of the item parameters. In the simulation, no such global change was assumed; all item parameters β_{ih}, $i = 1, \ldots, 20$, and effect parameters η_1 and η_2 are therefore estimable.)

The large sample size was used for the following reasons: consistency of the CML estimates in the LPCM can be taken for granted for theoretical reasons. Therefore, if the sample is large, we may expect that the estimates will be recovered very well. Should this not be the case, the results would indicate an essential failure of the estimation procedure. Actually, the β_{ih} were recovered with absolute deviations $|\beta_{ih} - \hat{\beta}_{ih}| \leq .20$ and a mean deviation of .09; the correlation between the β_{ih} and the $\hat{\beta}_{ih}$ was .998. The effect parameters were even recovered much more precisely, namely, $\hat{\eta}_1 = .97$ (for $\eta_1 = 1.00$) and $\hat{\eta}_2 = 2.02$ (for $\eta_2 = 2.00$). The superiority of the $\hat{\eta}_j$ over the $\hat{\beta}_{ih}$ is an expected phenomenon since the data contain much more statistical information on the treatment effect parameters than on any of the 79 item parameters β_{ih}. This shows that the LPCM can be used for measuring treatment effects even in cases where the item parameters (and, even more so, the individual subject parameters) are not estimated accurately.

B. *Run Time of the Program*

The CML estimation for the simulated data set required 4 mins. run time on an AT 386SX with 25 MHz. From the accuracy of the results and the speed of the iterative method we conclude that the present procedures are satisfactory and suitable for handling typical practical applications.

C. *Comparison of the BFGS with the Newton-Raphson Algorithm*

A comparison of the BFGS and the Newton-Raphson algorithms was carried out on another simulated data set under the more restrictive LRSM for which a published FORTRAN program using the Newton-Raphson method is available (Fischer & Parzer, 1991b). The LRSM estimates were practically identical under both programs (all absolute differences $\leq .0001$), which was as expected. The BFGS method was faster, though: 5 seconds for the BFGS versus 12 seconds for the Newton-Raphson method for 10 parameters estimated in that LRSM simulation. Therefore it is concluded that the BFGS method is more appropriate for LPCM applications where often more than 100 parameters have to be estimated; the size of the storage required, by the way, is the same as for the Newton-Raphson method.

D. *Small Sample Simulation*

One reviewer of a previous version of the paper asked for a simulation demonstrating the usefulness of the LPCM for realistic (i.e., much smaller) samples. Such a simulation was carried out, based on the design of the empirical example of section 6. We simulated 87 subjects with θ_v sampled from an $N(0, 1)$ according to our Model 1 (see section 6) comprising 119 parameters. The "true" β_{ih}, η_i, and τ_i were set to those values which had resulted as estimates in the empirical analysis. The differences between true parameters and the respective estimates had a standard deviation of 0.33, and the correlation resulted as 0.88. Upon repeating the simulation for Model 2 (see section 6), the "true" parameters $\eta = -.08$ and $\tau = -.23$ were estimated as $\hat{\eta} = -.09$ and $\hat{\tau} = -.23$, and the correlation of all parameters with their estimates was .95. Note that the reduction of the number of effect parameters in Model 2 as compared to Model 1 led to a considerable improvement of the precision. These results demonstrate that the parameters are well recovered even in small sample studies (if the model fits).

6. An Empirical Example

In the following we reanalyze the data of a clinical study on the effects of relaxation training (RT) on patients with certain psychosomatic disorders (called "functional syndrome"; Widowitz, 1987); the data have previously been used by Fischer and Parzer (1991a) for demonstrating an LRSM application.

The subjects were outpatients of the II. Internal Clinic of the University of Vienna who filled in a self-rating inventory (Zerssen, 1976) prior to a 7-week treatment period as well as one year later. The patients were randomly assigned to either of two groups: a treatment group that underwent relaxation training (RT Group) in addition to the usual medication, and a Control Group receiving only the medication. (For obvious reasons, medication was not withheld from any of the patients.) The design is described in Table 3. A total of 87 patients filled in the inventory; 37 of them belonged to the RT Group, and 50 to the Control Group. The inventory comprised (besides other self-rating scales) a set of 24 items related to typical "functional" symptoms, such as shortness of breath, faintness, difficulties of swallowing, fatigue, etcetera. The patients were asked to rate the actual intensity of each symptom on one 4-point rating scale.

Table 3.

Design of the empirical example (data of Widowitz, 1987): 24 real items, $I_1, \ldots, I_{24}$; two time points, T_1 and T_2; two treatment groups, G_1 and G_2; 72 V-items. V-items $I_1^*, \ldots, I_{24}^*$ are identical with real items $I_1, \ldots, I_{24}$. G_1 and G_2 respond to the same V-items at T_1, but to different V-items sets at T_2.

	V-items presented at	
	T_1	T_2
G_1	$I_1^*, \ldots, I_{24}^*$	$I_{25}^*, \ldots, I_{48}^*$
G_2	$I_1^*, \ldots, I_{24}^*$	$I_{49}^*, \ldots, I_{72}^*$

The effect of RT was the issue of main interest of the study. Fischer and Parzer (1991a) applied the LRSM to the 24 item scale of functional symptoms, whereby no significant RT effect could be inferred and—unfortunately—the LRSM had to be rejected for lack of fit. The authors concluded that the usage of the response categories was different in patients with a higher degree of impairment as compared to patients with only a slight impairment.

The 24 items scale will now be reanalyzed by means of the LPCM, which is more flexible owing to the larger number of item × category parameters β_{ih}. The first step is to formulate and test a general "quasi saturated" model where any interactions of treatments × (real) items × time points are allowed; this will be denoted Model 1. It is described by the following equations:

$$\delta_{vi1h} = \beta_{ih} \qquad \text{for all } S_v;\ \text{all } I_i;\ \text{time points } T_1;\ \text{V-items } I_l^*,\ l = 1, \ldots, 24; \tag{21}$$

$$\delta_{vi2h} = \beta_{ih} + h\tau_i \qquad \text{for } S_v \text{ of the Control Group, } G_1;\ \text{all } I_i;\ \text{time point } T_2;\ \text{V-items } I_l^*,\ l = 25, \ldots, 48; \tag{22}$$

$$\delta_{vi2h} = \beta_{ih} + h(\tau_i + \eta_i) \qquad \text{for } S_v \text{ of the RT group, } G_2;\ \text{all } I_i;\ \text{time point } T_2;\ \text{V-items } I_l^*,\ l = 49, \ldots, 72. \tag{23}$$

Therein,

β_{ih} are the (real) item parameters (as before),
τ_i is the effect of medication in regard to item (symptom) I_i, and
η_i is the effect of RT in regard to item (symptom) I_i.

On the one hand, this model makes allowance for differential effects of the treatments on each item (or symptom; more precisely: on the latent tendency to produce that particular symptom); on the other hand, it assumes generalizability of treatment effects over patients and response categories. The effect parameter estimates for Model 1 are given in Table 4. As is seen, these parameter estimates have an average precision comparable to the one found in the small sample simulation of section 5 above. Almost all effects are negative, indicating that there is some improvement of the state of the

Table 4.

Treatment effect parameter estimates, $\hat{\tau}_i$ and $\hat{\eta}_i$, and their respective standard deviations for the Widowitz (1987) data.

i	$\hat{\tau}_i$	$\hat{\sigma}_{\tau_i}$	i	$\hat{\eta}_i$	$\hat{\sigma}_{\eta_i}$	i	$\hat{\tau}_i$	$\hat{\sigma}_{\tau_i}$	i	$\hat{\eta}_i$	$\hat{\sigma}_{\eta_i}$
1	−0.41	0.17	1	−0.32	0.15	13	−0.57	0.41	13	0.27	0.22
2	−0.10	0.13	2	−0.31	0.13	14	−0.23	0.12	14	−0.30	0.12
3	−0.38	0.15	3	−0.58	0.15	15	−0.32	0.16	15	−0.36	0.15
4	−0.18	0.12	4	−0.48	0.12	16	−0.10	0.18	16	−0.57	0.20
5	−0.18	0.15	5	−0.35	0.14	17	−0.49	0.15	17	−0.41	0.15
6	−0.23	0.12	6	0.00	0.12	18	−0.63	0.19	18	−0.41	0.16
7	−0.11	0.12	7	−0.31	0.11	19	−0.30	0.13	19	−0.03	0.12
8	−0.26	0.12	8	0.04	0.11	20	−0.74	.049	20	−0.01	0.30
9	−0.05	0.12	9	0.01	0.11	21	−1.10	0.38	21	−0.12	0.30
10	−0.19	0.12	10	−0.11	0.13	22	−0.64	0.13	22	−0.42	0.12
11	−0.45	0.15	11	0.05	0.13	23	−0.34	0.14	23	−0.21	0.12
12	0.01	0.13	12	−0.24	0.12	24	−0.74	0.12	24	−0.41	0.12

patients, that is, the latent tendencies of producing the symptoms decrease. (We refrain from presenting the item parameter estimates since they are of little interest here anyway.) The log likelihood is $\ln L_1 = -4154.81$. This model will be used as the H_1 for testing several null-hypotheses about change.

One such hypothesis is that the treatment effects, τ_i and η_i, *generalize over items* (symptoms); this H_0, formalized within the model by setting $\tau_i = \tau$ and $\eta_i = \eta$ for all I_i, will be denoted Model 2. The log-likelihood under Model 2 results as $\ln L_2 = -4181.38$, hence the likelihood ratio statistic for H_0 against H_1 is

$$\chi^2 = -2(\ln L_2 - \ln L_1) = -2(-4181.38 + 4154.81) = 53.14,$$

with df $= 2 \times 24 - 2 = 46$ and $\chi^2_{.99} = 71.20$ (the difference in the numbers of effect parameters); this is nonsignificant. Hence there is no indication of a differential effect of the treatments in regard to particular symptoms, so that Model 2 can be accepted. The estimates of the effect parameters and of their standard deviations under Model 2 are

$$\hat{\tau} = -.23, \qquad \hat{\sigma}_\tau = .045,$$

$$\hat{\eta} = -.08, \qquad \hat{\sigma}_\eta = .071.$$

This shows that the medication effect is significant, but that there is no significant RT effect. (Obviously, the decrease in the number of effect parameters compared to Model 1 again improves the precision of the estimates).

The central question of the study of Widowitz (1987) was whether—and if at all, in what way—RT does support the medical treatment. This can be answered more precisely and independently of the existence/nonexistence of a medication effect τ by means of a conditional likelihood ratio test of the null-hypothesis H_0: $\eta = 0$, denoted

Model 3, against Model 2 as the alternative. The log-likelihood under H_0 is $\ln L_3 = -4181.99$, and hence

$$\chi^2 = -2(\ln L_3 - \ln L_2) = -2(-4181.99 + 4181.38) = 1.22,$$

with df = 1. This is nonsignificant even at $\alpha = .05$, $\chi^2_{.95} = 3.84$. It is therefore not possible to say that RT had a palliative effect on the functional syndrome. (This conclusion is also in accordance with the respective hypothesis test in Fischer & Parzer, 1991a, based on the more restrictive LRSM).

Another way of testing the H_0 of no RT effect is to set $\eta_i = 0$ for all I_i in (23), denoted Model 3*, and to compare this with Model 1: the log-likelihood under Model 3* is $\ln L_{3*} = -4172.69$, and the test statistic

$$\chi^2 = -2(\ln L_{3*} - \ln L_1) = -2(-4172.69 + 4154.81) = 35.76,$$

with df = 24, which again is nonsignificant (even at $\alpha = .05$ $\chi^2_{.95} = 36.41$).

Medication, on the other hand, did have a significant effect: Model 4, assuming no change at all, that is, H_0: $\tau_i = \eta_i = 0$ for all I_i, yields $\ln L_4 = -4209.13$. Comparing Model 4 to Model 1 results in the statistic

$$\chi^2 = -2(\ln L_4 - \ln L_1) = -2(-4209.13 + 4154.81) = 108.64,$$

with df = $2 \times 24 = 48$ and $\chi^2_{.99} = 76.20$, which is significant. We conclude that change occurred, attributable to a medication effect. (There is no way of controlling for possible spontaneous remissions though, since the design contained no untreated control group.)

The inevitable drop of bitterness in this empirical example is the test of fit that has to be made for the validity of Model 1 underlying the above analysis. Such tests can be made by splitting the patient samples according to (a) score (= degree of impairment), (b) gender, and (c) education. Each of these criteria is potentially influential on the effectiveness of the treatments. Estimating all parameters of Model 1 separately for patients with above-average versus below-average scores yields the log likelihoods $\ln L_1^{(H)} = -2145.83$ and $\ln L_1^{(L)} = -1903.12$, denoted Model 5, such that the conditional likelihood ratio statistic for Model 1 against Model 5 is

$$\begin{aligned}\chi^2 &= -2(\ln L_1 - \ln L_1^{(H)} - \ln L_1^{(L)}) \\ &= -2(-4154.81 + 2145.83 + 1903.12) = 211.72.\end{aligned}$$

The degrees of freedom are equal to the number of independent parameters under Model 5 minus the corresponding number under Model 1: there are $3 \times 24 - 1 = 71$ item parameters and $2 \times 24 = 48$ treatment effect parameters under Model 1, yielding a total of 119 independent parameters, and there are twice as many under Model 5; hence, df = 119. For $\alpha = .01$ the critical value is $\chi^2_{.99} = 157.80$, implying that the test is significant. Similarly, for gender, $\ln L_1^{(M)} = -734.77$ for males, and $\ln L_1^{(F)} = -3329.22$ for females, yielding

$$\chi^2 = -2(-4154.81 + 734.77 + 3329.22) = 181.64,$$

with the same df and critical value as before, which is significant. Finally, for education, $\ln L_1^{(E)} = -1205.96$ for patients with a higher education, and $\ln L_2^{(U)} = -2881.67$ for those without, giving

$$\chi^2 = -2(-4154.81 + 1205.96 + 2881.67) = 134.36,$$

again with the same df and critical value, which is nonsignificant. Hence, we see that Model 9 has to be rejected on the basis of two of the three criteria. Looking at the size of the χ^2 statistics we venture the interpretation that patients with different degrees of impairment use the response categories differently—as was also conjectured by Fischer and Parzer (1991a).

So we hypothesize that response style influences the response behavior of the patients such that patients with relatively little complaints have a particular preference for response category C_0 ("no complaint"), and conversely, patients with very marked complaints emphasize them by preferring category C_3 ("strong complaint"). This hypothesis is modeled by introducing a response style parameter λ as follows: For the patients with a medium overall complaints score, Model 2 is presumed; for those patients who are low on that score, β_{i0} is replaced by $\beta_{i0} + \lambda$ or, in order to preserve the normalization of the δ_{lh}, β_{ih} is replaced by $\beta_{ih} - \lambda$ for $h = 1, 2, 3$. Conversely, for the high-scoring subgroup, β_{i3} is replaced by $\beta_{i3} + \lambda$. This hypothesis is tested by splitting the sample in three groups, patients with $r_v < 40$, $40 < r_v < 60$, and $r_v > 60$, each group comprising 29 subjects, and comparing the log-likelihood under Model 2 (as H_0) with that of Model 6 (as H_1):

$$\chi^2 = -2(\ln L_2 - \ln L_6) = -2(-4181.38 + 4179.73) = 3.30,$$

with df = 1, which is barely nonsignificant at $\alpha = .05$, $\chi^2_{.95} = 3.84$. The lack of significance is most probably due to the very small subsamples (29 patients each). Therefore, we conjecture again that response style ("extremity") biases patients' responses to the inventory.

For the sake of brevity we refrain from presenting more hypothesis tests. In particular, it would be possible to introduce a more general (and considerably more complicated) saturated model as the basis for further hypothesis testing. For the present illustration, however, it suffices to demonstrate the feasibility of tests of relevant hypotheses in the face of a practical research problem.

7. Discussion

One particularly attractive feature of the LPCM is its greater flexibility as compared to the LRSM. The LPCM is therefore useful for analyzing data sets when the simpler (and easier to interpret) LRSM fails. However, it must be borne in mind that any gain in flexibility of a model must be paid for with a loss of simplicity and, consequently, of ease of interpretation. In that respect the LPCM is a compromise between the scalar parameterization of items and subjects in the LRSM and the vectorial parameterization in the polytomous Rasch model (and its applications to measurement of change, see Fischer, 1972, 1977).

The relatively large number of parameters that have to be estimated in the LPCM requires special attention to the numerical procedures needed for the conditional maximum likelihood approach. As we have seen, the recurrent computation of the γ-functions and the BFGS (quasi-Newton) approach to the CML equations are very effective for solving the numerical problems involved. Both the simulation and the empirical example demonstrate that these methods are applicable to typical research problems arising in applied psychology. The observed stability of the BFGS method as applied to the LPCM rests on the convenient properties (convexity of the log-likelihood function) of the exponential family defined by the LPCM.

Another important property of the LPCM as applied to repeated measurement designs is that the effect parameters can be estimated much more precisely than the item parameters. Hence we may summarize our results by stating that the LPCM is a

useful and practical tool for analyzing polytomous response data, particularly so for testing hypotheses on change.

References

Andersen, E. B. (1983). A general latent structure model for contingency table data. In H. Wainer & S. Messik (Eds.), *Principals of modern psychological measurement* (pp. 117–138). Hillsdale/NJ: Erlbaum.

Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*, 581–594.

Armijo, L. (1966). Minimization of functions having Lipschitz-continuous first partial derivatives. *Pacific Journal of Mathematics, 16*, 1–3.

Churchhouse, R. F. (1981). *Handbook of applicable mathematics, Vol. III*. Chichester—New York: J. Wiley.

Fischer, G. H. (1972). A measurement model for the effect of mass-media. *Acta Psychologica, 36*, 207–220.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [Introduction to mental test theory]. Bern: Huber. (In German)

Fischer, G. H. (1977). Some probabilistic models for the description of attitudinal and behavioral changes under the influence of mass communication. In W. F. Kempf & B. Repp (Eds.), *Mathematical models for social psychology* (pp. 102–151). Bern: Huber, and New York: Wiley.

Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika, 54*, 599–624.

Fischer, G. H. (1991). A new methodology for the assessment of treatment effects. *Evaluación Psicológica/ Psychological Assessment, 7*, 117–147.

Fischer, G. H., & Allerup, P. (1968). Rechentechnische Fragen zu Raschs eindimensionalem Modell [Numerical problems in the unidimensional model of Rasch]. In G. H. Fischer (Ed.), *Psychologische Testtheorie* [Psychological test theory] (pp. 269–280). Bern: Huber. (In German)

Fischer, G. H., & Parzer, P. (1991a). An extension of the rating scale model with an application to the measurement of treatment effects. *Psychometrika, 56*, 637–651.

Fischer, G. H., & Parzer, P. (1991b). LRSM: Parameter estimation for the linear rating scale model. *Applied Psychological Measurement, 15*, 138.

Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models*. (Doctoral dissertation at the University of Twente, Department of Education). The Hague: CIP-Gegevens Koninklijke Bibliotheek.

Glas, C. A. W., & Verhelst, N. D. (1989). Extensions of the partial credit model. *Psychometrika, 54*, 635–659.

Gustafsson, J.-E. (1980). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement, 40*, 377–385.

Jacobsen, M. (1989). Existence and unicity of MLEs in discrete exponential family distributions. *Scandinavian Journal of Statistics, 16*, 335–349.

Kosmol, P. (1989). *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. Stuttgart: Teubner.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the IV. Berkeley symposium on mathematical statistics and probability, Vol. 4* (pp. 321–333). Berkeley: University of California Press.

Rasch, G. (1965). *Statistisk seminar* [Statistical seminar]. Cogenhague: University of Copenhague, Department of Mathematical Statistics. (Notes taken by J. Stene)

Rasch, G. (1967). An informal report on a theory of objectivity in comparisons. In L. J. Th. van der Kamp & C. A. J. Vlek (Eds.), *Psychological measurement theory* (pp. 1–19). Leyden: University of Leyden. (Proceedings of the NUFFIC international summer session in science in "Het Oude Hof", The Hague, July 14–28, 1966)

Widowitz, E. (1987). *Der Effekt autogenen Trainings bei funktionellen Erkrankungen*. [The effect of "autogenous training" on the functional syndrome]. Unpublished master's thesis. Vienna: University Vienna, Department of Psychology.

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

Zerssen, D. v. (1976). *Klinische Selbstbeurteilungsskalen (KSb-S) aus dem Münchner Psychiatrischen Informations-System (PSYCHIS München)* [The clinical self-rating scales (KSb-S) of the "Munich Psychiatric Information System" (PSYCHIS)]. Weinheim: Beltz.

Manuscript received 7/8/92
Final version received 3/23/93