

# Some Applications of Logistic Latent Trait Models with Linear Constraints on the Parameters

Gerhard H. Fischer and Anton K. Formann  
University of Vienna

The linear logistic test model (LLTM), a Rasch model with linear constraints on the item parameters, is described. Three methods of parameter estimation are dealt with, giving special consideration to the conditional maximum likelihood approach, which provides a basis for the testing of structural hypotheses regarding item difficulty. Standard areas of application of the LLTM are surveyed, including many references to empirical studies in

item analysis, item bias, and test construction; and a novel type of application to response-contingent dynamic processes is presented. Finally, the linear logistic model with relaxed assumptions (LLRA) for measuring change is introduced as a special case of an LLTM; it allows the characterization of individuals in a multidimensional latent space and the testing of hypotheses regarding effects of treatments.

The widespread use of tests as entrance examinations for schools and universities has greatly stimulated the further development of test theory in the U.S.A. and has given rise to innovative technologies of testing based on item response models. In these applications the three-parameter logistic model has been preferred because its item characteristic curves (ICCs) can be flexibly fitted to a variety of items and item formats in use. In most European countries, on the contrary, testing has so far been of only secondary importance in admission to schools or selection of curricula. Testing has been limited mainly to the traditional domain of psychological diagnostics for the individual. As a consequence, there has been less interest in new testing technologies than in the theoretical foundation of tests as instruments of psychological research. Latent trait theory in Europe has developed more towards a theory of psychological measurement and towards modeling the process of problem solving, in which the one-parameter logistic (Rasch) model has played an outstanding role. These research aims have suggested going beyond the description of response probabilities in terms of item and person parameters, tracing the former back to the influences of the cognitive operations involved and the latter, at least in part, to learning processes.

The present paper deals with a class of linearly restricted logistic models which can be viewed as formalizations of structural hypotheses regarding the psychological complexity of test items and which may serve for testing such hypotheses empirically. The most important formal results, a survey of typical areas of application, and an empirical example of a novel type of application will be given.

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 6, No. 4, Fall 1982, pp. 397-416  
© Copyright 1982 Applied Psychological Measurement Inc.  
0416-6216/82/040397-20\$2.00

### The Rasch Model and Its Generalizations

The latent trait model introduced by Rasch (1960), called the Rasch model (RM), explains the probability of a correct response of person  $a$  to item  $i$  as the function of one ability  $\theta_a$  (or  $\xi_a$ , respectively) of the individual and of the difficulty  $b_i$  (or easiness  $\varepsilon_i$ ) of the item:

$$P(+|i, a) = P_{ia} = \frac{\exp(\theta_a - b_i)}{1 + \exp(\theta_a - b_i)} = \frac{\xi_a \varepsilon_i}{1 + \xi_a \varepsilon_i} \quad [1]$$

for  $a = 1, \dots, N$  (individuals) and  $i = 1, \dots, n$  (items), with  $\xi_a = \exp(\theta_a)$  and  $\varepsilon_i = \exp(-b_i)$ . Both these parameterizations will be needed below. It is not possible to deal here with the question whether and under which conditions the parameters of the RM can be uniquely estimated. (See recent articles on this topic by Andersen, 1980; Fischer, 1981; Gustafsson, 1980a; Wainer, Morgan, & Gustafsson, 1980). However, it is well known and has been especially stressed by Rasch (1960, 1967) that the item parameters can be estimated independently of the true values of the person parameters; therefore, the term "sample-free" has come into use in connection with the RM.

### The Rasch Model as Conjoint Measurement

The RM can also be considered a method of conjoint measurement of item and person parameters based on the logit-transformed reaction probabilities in a two-way factorial persons  $\times$  items design,

$$\text{logit}(P_{ia}) = \theta_a - b_i. \quad [2]$$

The parameters  $\theta_a$  and  $b_i$  can be interpreted as main effects of persons and items, respectively. The separability of both sets of parameters from each other is a consequence of the simple additive concatenation in Equation 2. (For further details on this see Rasch, 1968, 1977.) The close analogy of Equation 2 to an analysis of variance without interactions has suggested a generalization of this additive model to more than two factors:

$$\text{logit}(P_{aij\ell\dots k}) = \theta_a - b_i + \alpha_j + \beta_\ell + \dots + \omega_k, \quad [3]$$

where, for example, the parameters  $\alpha_j$  could be the effects of different points of time of observation between which learning may have occurred, the  $\beta_\ell$ 's those of different experimental conditions, and, finally, the  $\omega_k$ 's the influences of different modes of test presentation. The above-mentioned separability of the parameters implies that the parameters of the levels of each single factor can be estimated separately and independently of all the other parameters. Such analyses of multifactorial complete designs have already been suggested by Micko (1969, 1970), Falmagne (quoted from Micko, 1970), Scheiblechner (1971) and Kempf (1972); see also the more detailed discussion in Fischer (1974, pp. 340-353). Since, however, data for this type of complete multifactorial design are rare, the practical use of this approach is rather limited.

### The Linear Logistic Test Model

A generalization proposed by Scheiblechner (1972) that has a much broader spectrum of applications is the so-called linear logistic test model (LLTM); it consists of the RM defined in Equation 1 along with the linear restrictions

$$b_i = \sum_j q_{ij} \eta_j + c, \quad j = 1, \dots, m, \quad m < n. \quad [4]$$

Therein the  $\eta_j$ 's denote the *basic parameters* of the model and the  $q_{ij}$ 's the weights of the  $\eta_j$ 's for each of the item parameters  $b_i$ . Depending on the type of application, the basic parameters can be interpreted either as the contribution of hypothetical cognitive operations to the total difficulty of the items, with the  $q_{ij}$ 's being the minimum number of times operation  $j$  has to be used for solving item  $i$ , or as the effect parameters of the factor levels in a multifactorial design with design matrix  $Q$ . Last but not least, the  $\eta_j$ 's can also be used to describe processes of change in the individuals, as, for example, caused by treatments or by practice. How hypotheses can be formulated by means of  $Q$  will be demonstrated in the applications below. The constant  $c$  is the usual normalization constant of the RM. (The  $b_i$ 's are only defined up to an additive constant).

**Parameter Estimation**

For estimating the basic parameters, several methods are known, one of which stands out because of its theoretical advantages, namely, the conditional maximum likelihood (CML) method. As is well known, the conditional likelihood function of a data matrix  $U$ , given the raw scores of the individuals, is independent of the parameters  $\theta_a$ . As applications of the model to cases of incomplete data will be discussed below, the conditional likelihood function for incomplete data is given here:

$$L(U | r_1, \dots, r_N; V) = \frac{\prod_i \epsilon_i^{s_i}}{\prod_a \gamma_{r_a}}, \quad [5]$$

where  $U = [u_{ia}]$  denotes the usual item-score matrix; missing observations are formally represented by some arbitrary value  $u$ ,  $0 < u < 1$ . Further,  $V = [v_{ia}]$  is a matrix with elements  $v_{ia} = 1$  if some response of person  $a$  to item  $i$  has been observed, and  $v_{ia} = 0$  otherwise;  $r_a = \sum_i u_{ia} v_{ia}$  is the raw score of individual  $a$  and  $s_i = \sum_a u_{ia} v_{ia}$  the marginal sum of item  $i$ ; finally,  $\gamma_{r_a}$  denotes the elementary symmetric function of order  $r_a$  of the variables  $\epsilon_1 v_{1a}, \dots, \epsilon_n v_{na}$ . (For further details see Fischer, 1981.) It should be stressed that the missing observations may not be responses to subjectively difficult items omitted by the testees but should be due to design. (For the problem of "omits," see Lord, 1980, pp. 226-230).

CML estimates of the basic parameters of the LLTM are obtained by taking the logarithm of the likelihood defined in Equation 5 and then differentiating with respect to the  $\eta_j$ 's. The estimation equations are

$$\sum_i q_{ik} \left[ s_i - \sum_a \frac{\epsilon_i v_{ia} \gamma_{r_a}^{(i)}}{\gamma_{r_a}} - 1 \right] = 0, \quad k = 1, \dots, m. \quad [6]$$

The functions  $\gamma_{r_a-1}$  in Equation 6 denote, analogously to  $\gamma_{r_a}$ , the elementary symmetric functions of order  $r_a-1$  with the arguments  $\epsilon_1 v_{1a}, \dots, \epsilon_{i-1} v_{i-1,a}, \epsilon_{i+1} v_{i+1,a}, \dots, \epsilon_n v_{na}$  (i.e., the  $i^{th}$  parameter  $\epsilon_i$  is omitted). The derivation of these equations can be found in Fischer (1973; 1974, pp. 353-359). A FORTRAN program using a gradient method has been published by Fischer and Formann (1972; see also Fischer, 1974, pp. 531-554). Its application is, in general, limited to about 40 items because obtaining

numerical accuracy of the functions  $\gamma_{ra}$  and  $\gamma_{r_{a-1}}$  becomes increasingly difficult when  $n$  becomes large. More refined recursion formulae for the elementary symmetric functions admit up to about 80 items (see Gustafsson, 1980a). In addition, new program versions use the Newton-Raphson method by which a much faster convergence is attained (about five iterations suffice in most cases). The conditions which the data  $U$  and the structural matrix  $Q$  must jointly fulfill so that the Equations 6 possess a unique solution have been derived by Fischer (1982). Since these are rather complicated, they cannot be stated here. However, in summary, it can be said that the CML method, having a range of up to 80 items and fast convergence, is suited for most cases in practice.

In the unlikely event (1) that there are more than 80 items which conform to the RM and to which the LLTM can be applied meaningfully and (2) that at least some of the testees have responded to more than 80 items, a simple alternative for parameter estimation is to maximize the unconditional likelihood function

$$L(U|V) = \prod_i \prod_a \left[ \frac{(\epsilon_i \xi_a)^{u_{ia}}}{1 + \epsilon_i \xi_a} \right]^{v_{ia}} \quad \text{with } \epsilon_i = \exp(-\sum_j q_{ij} \eta_j - c) . \quad [7]$$

The estimation equations resulting for this unconditional maximum likelihood (UML) method are

$$\sum_i q_{ik} (s_i - \sum_a P_{ia} v_{ia}) = 0 \quad (k = 1, \dots, m) \quad [8]$$

for the basic parameters (with given person parameters  $\theta_a$ ) and

$$r_a = \sum_i P_{ia} v_{ia} \quad (a = 1, \dots, N) \quad [9]$$

for the person parameters (with given person parameters  $b_i$ ). Assuming approximate values for the person parameters, the system of Equations 8 can be solved for the parameters  $\eta_1, \dots, \eta_m$  by means of the Newton-Raphson method; then these solutions can be inserted into Equation 9 in order to solve each of these equations, one for each vector  $(v_{1a}, \dots, v_{na})$  and each raw score  $r_a$ , for the respective person parameter. These are again substituted into Equation 8, and so forth. If the number of item parameters is not too small, the bias of this UML method will be rather slight (cf. Andersen, 1973; 1980, p. 244; Haberman, 1977).

However, this simple UML approach has one serious drawback: The CML method allows the computation of the conditional likelihood as a function of the  $\eta_j$ 's and the direct comparison of this with the conditional likelihood function of the RM. The former is obtained by inserting the  $\hat{\eta}_j$ 's into Equation 5; let the result be denoted by  $L(\hat{\eta}_1, \dots, \hat{\eta}_m)$ . The latter is obtained by estimating the item parameters of the RM and inserting these  $\hat{b}_i^R$ 's into Equation 5; let the resulting likelihood function be denoted by  $L(\hat{b}_1^R, \dots, \hat{b}_n^R)$ . Then, the following likelihood ratio (LR) statistic can be obtained,

$$\chi^2 = -2 \ln [L(\hat{\eta}_1, \dots, \hat{\eta}_m) / L(\hat{b}_1^R, \dots, \hat{b}_n^R)] \quad [10]$$

which is asymptotically  $\chi^2$ -distributed with  $df = n - 1 - m$  as long as the LLTM holds. Hence, the validity of Equation 4 can be tested in an elegantly simple manner, and, all the more so, as the likelihood functions needed result as a by-product of parameter estimation. Such LR tests, however, are not known for the UML approach, and it is not likely that they can easily be established because the unconditional likelihood function depends on the person parameters  $\theta_a$ , the number of which tends to

infinity as  $N \rightarrow \infty$ . For this reason, the CML method has been used in almost all applications of the LLTM.

Besides testing the LLTM (as an  $H_0$ ) against the RM (as  $H_1$ ), other LR tests are also important for practical applications, for example, to test whether the basic parameters  $\eta_i$  are the same in specified subsamples of individuals or of items. Let the two samples to be compared be denoted by  $I$  and  $II$  and the total sample by  $T$ , and let the CML estimates be computed separately for  $I$ , for  $II$ , and for the total sample  $T$ ; then the asymptotic LR test

$$\chi^2 = -2 \ln \left[ \frac{L(T)}{L(I)L(II)} \right] \tag{11}$$

holds with  $L$  defined again as in Equation 5 and with  $df$  equal to the difference of the numbers of independent parameters in the denominator and in the numerator of Equation 11. This obviously allows the testing of many more types of substantive hypotheses. Furthermore, constraints with different numbers of parameters in Equation 4 can be tested against each other. The main interest in applying the LLTM is thus for testing hypotheses rather than for merely estimating the parameters.

An interesting alternative to the CML and UML approaches is the EM algorithm of Dempster, Laird, and Rubin (1977), which has recently been adopted for latent trait theory by Bock and Aitkin (1981) and by Thissen (1982). It requires the specification of some distribution of the  $\theta_a$ 's, e.g., a normal probability density function (*pdf*). Although a first comparison of CML and EM estimates has indicated a good correspondence (Thissen, 1982), the main problem with this method again seems to be the testing of hypotheses: Suppose it is desired to test the  $H_0$  of equal cognitive complexity (i.e., equal basic parameters for the cognitive operations) in two different populations (e.g., males vs. females); then the assumption of two different normal *pdf*'s in these two groups precludes the existence of one common normal *pdf* for the total sample. Therefore, LR tests as described above would not be feasible.

### Typical Applications of the LLTM

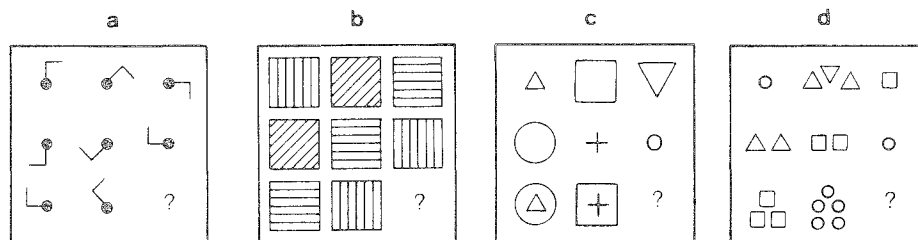
The early applications of the LLTM primarily attempted to explain item difficulty on the basis of their hypothetical cognitive structures. Scheiblechner (1972) studied the cognitive complexity of certain propositions of formal logic which were, however, presented in concrete graphical form. It turned out to be possible to explain item difficulty by means of only three operations ("negation," "disjunction," and "considering asymmetry"). Spada, Fischer, and Heyner (1974) analyzed tasks taken from elementary mechanics; and Fischer (1973) used mathematical problems taught in high school. These first attempts showed that the LLTM is a useful tool for item analysis but that a good fit of the model is difficult to attain or, rather, that it is attained only if the item material was constructed carefully and deliberately, with an eye to the model.

With this in mind, Formann (1973; see also Formann & Piswanger, 1979) undertook the construction of a new nonverbal intelligence test similar to Raven's Progressive Matrices. While retaining the item format, he systematically designed the items according to predetermined construction rules. Three factors (or facets) were assumed to be relevant for item difficulty: One factor with the levels "continuation," "variation," and "superimposition" defined the *rule* that had to be applied by the subjects; another factor with levels "horizontal," "vertical," and "horizontal and vertical" defined the *direction* for applying the rule; the third factor with levels "form," "pattern," "number," and

"array" characterized the aspect of the *graphic components* to which the rule referred. Figure 1 illustrates the meaning of these three factors. Since the graphic components can be realized by means of a practically unlimited number of geometric representations ("form" can be, e.g., a circle, a triangle, a square), and since one or several rules can occur simultaneously in an item, the three factors generate a universe of items which may either be basically different because of their structures or, if they have the same structure, superficially seem different to the testees because of the graphical components.

Formann's empirical investigation of a sample of 42 of these items (with two rules each) by means of the RM proved that these construction rules enabled him to design items with very different difficulty parameters. However, nine items had to be deleted because they did not fit the RM. For the other items, several hypotheses with different numbers of cognitive operations, even considering interactions between the three factors, were tested using the LLTM and comparing its likelihood with that of the RM (as  $H_1$ ). The attempt to explain item difficulty as a function of item structure was only partially successful, since all the LR tests turned out to be significant. That  $H_0$  which assumed the independence of the three factors, and could thus be interpreted most readily, was regarded as the most acceptable approximation to the more complex reality. According to this  $H_0$ , the difficulty of each item is an additive combination of the parameters of the respective rules, of those of the required directions, and of those of the relevant properties, independent of the specific combinations of the factor levels of the three facets rule, direction, and relevant property.

**Figure 1**  
The Rules for Constructing the Matrices Items Illustrated by Four Items,  
Three Consisting of One Component—(a) Continuation, Array,  
Horizontal, (b) Variation, Pattern, Horizontal & Vertical,  
(c) Superimposition, Form, Vertical—and One Comprising Two Components—  
(d) Variation, Form, Horizontal & Vertical, and Superimposition, Number, Vertical



Within the limits of empirical validity of the LLTM, the difficulty of new items which have never been presented to anyone can be predicted or tasks of prespecified difficulty can be constructed, using the parameter estimates of the elementary operations of the respective test materials. This new possibility offered by the LLTM seems to be of special interest in view of individualized adaptive testing. There, large numbers of—it is to be hoped— unidimensional items, which ought to vary within a broad range of difficulty, are needed so that for each individual the optimal choice, depending on his/her ability level, is possible. The use of the described universe of items for a branched testing procedure was tried out by Fischer and Pendl (1980); their approach attempted to combine the advantages of simple paper-and-pencil administration with that of the increased precision of adaptive testing.

A necessary requirement for such uses, as well as for constructing parallel tests, is the constancy of the parameters of the cognitive operations within a defined population of persons. This can be as-

sumed to hold, by and large, for the described matrices test, as can be concluded from later investigations of Piswanger (1975; see also below), Nährer (1977), and Habon (1981).

Scheiblechner (1972) attempted to adjust the LLTM to take account of the phenomenon that those items which are presented at the end of a test are relatively easier than items with a similar structure presented at the beginning. Such systematic deviations can be attributed to learning which occurs during test taking, independent of whether the items are solved or not. (In contrast to this assumption, Kempf, 1977, formulated a response-contingent dynamic test model).

Two types of such learning processes can be distinguished, namely, *global vs. operation learning*. In global learning it is assumed that all item parameters  $b_i$  can be decomposed into the operation parameters  $\eta_j$  and one additional parameter,  $\delta$ , which measures the effect of learning derived from practicing the operations. The weight assigned to this parameter  $\delta$  denotes the total number of operations contained in the items 1,2, ...,  $i - 1$ , i.e., how often it was already possible to practice operations, independent of which concrete operations these actually were. Equation 4 then becomes

$$b_i = \sum_{j=1}^m q_{ij} \eta_j - \delta \sum_{\ell=1}^{i-1} \sum_{j=1}^m q_{\ell j} + c . \tag{12}$$

If, on the other hand, learning occurs specifically for each of the operations (operation learning), as many learning parameters  $\delta_j$  are added as there are  $\eta_j$ 's, namely, one for each operation. The respective weights specify how often operation  $j$  was contained in the preceding  $i - 1$  items and how often, therefore, it was possible to practice it. Then, instead of Equation 12

$$b_i = \sum_{j=1}^m q_{ij} (\eta_j - \delta_j \sum_{\ell=1}^{i-1} q_{\ell j}) + c . \tag{13}$$

is obtained.

The obvious argument that practicing an easy operation is not likely to facilitate mastering some difficult operation, which perhaps has not even occurred in the test so far, speaks against the hypothesis of global learning. In this respect, the assumption of Equation 13 appears to be more plausible. On the other hand, Equation 13 implies that the difficulty of an operation that has been sufficiently practiced will become negative in the long run. This is a psychologically meaningless consequence because then the involvement of this operation in solving an item would make the item easier. Spada (1976, pp. 146-151; 1977) has dealt with this problem extensively and has made suggestions on how to overcome it. He investigated tasks from elementary mechanics and concluded that the LLTM was superior to the other models applied by him, namely, the deterministic model of Scandura (1973), the probabilistic automata model of Suppes (1977), and the logistic automata model formulated by Spada, even if the LLTM fitted only approximately. But the LLTM yielded interpretable parameters for his three sets of items, and the results were corroborated by cross-validations carried out in addition to the usual tests of fit.

The possibility of detecting learning processes within the test-taking situation is a methodological novelty, not to speak of the other assets of the LLTM. This becomes very relevant in connection with adaptive testing because this assumes the availability of a homogeneous pool of items with parameters which are constant *independent of the order of presentation*. This problem of learning seems to have been ignored in the literature on adaptive testing. Applications of the LLTM, however, have pointed very definitely to the existence of such changes in the item parameters which might be a reason for questioning the feasibility of adaptive testing. If, however, one of the above-mentioned models

for learning (or similar ones) proved to be sufficiently valid, then the order effects of item presentation could be taken into account in estimating the person parameters.

Finally, some more applications of the LLTM should be mentioned: Haudek (1973) studied number sequences; Heinrich (1975), syllogistic reasoning; and Kubinger (1979, 1980), the difficulties experienced by psychology freshmen when undertaking statistical analyses of experimental data.

#### Item Bias

Within the framework of latent trait theory, an item is called biased if a person's probability of a correct response, given the ability parameter  $\theta_a$ , depends on the membership of that individual in some group of persons, e.g., an ethnic group (see Lord, 1980, pp. 217-224; Nenty & Dinero, 1981; Scheuneman, 1975, 1980). Hence, the empirical analysis of item bias is closely related to testing the fit of the model to the data. The presence of item bias precludes the validity of the latent trait model for the total population of individuals.

The RM is particularly suited for studying item bias, since the item parameters can be estimated sample-free if the model is empirically valid. Deviations in the estimates in different groups of persons point to lack of validity of the model; parameter differences can be interpreted as item bias, at least as long as the model is true within each of the separate groups.

Even in the early applications of the RM, generalizability of the item parameter estimates across populations was seen as an important desideratum (e.g., Kearney, 1966; Mellenbergh, 1972). Spada (1969, 1970) examined some scales of the "LPS" intelligence test (Horn, 1962) whereby the total sample was in turn split according to gender, age, and total raw score, enabling comparisons of the resulting parameter estimates. Among other results, he detected some items that were easier for males, whereas others were easier for females. In these comparisons, he used graphical controls and an approximate  $\chi^2$ -statistic proposed by Fischer and Scheiblechner (1970) for detecting differences between item parameter estimates in two subsamples.

A multitude of methods for the statistical analysis of item bias in two or more subsamples by means of the RM exists: Wright and Panchapakesan (1969) and Wright, Mead, and Draba (1976) suggested analyzing the residuals, that is, the differences between observed responses  $u_{ia}$  and their expected values  $P_{ia}$ . Andersen (1973; 1980, pp. 251-258) derived an asymptotic LR test for the comparison of groups (in standard applications of the RM, usually score groups). His test statistic is a global measure for group differences, but not for the bias of single items. Remarkable for its simplicity is a procedure by Scheuneman (1975, 1980) that holds for any single item: Within the group of individuals with raw score  $r_a$ , the probability of a correct response to item  $i$  is

$$P(u_{ia} = 1 | r_a, \theta_a, \epsilon_1, \dots, \epsilon_n) = \epsilon_i \gamma_{r_a - 1}^{(i)} / \gamma_{r_a} \quad [14]$$

and therefore does not depend on  $\theta_a$  (as far as the RM holds). In a contingency table with score groups as rows, (ethnic) groups as columns, and frequencies of correct responses to item  $i$  as cell entries, the  $H_0$  of the absence of item bias can be tested by comparing observed and expected frequencies with a  $\chi^2$  test.

Recently, copious literature on tests of fit for the RM and on their sensitivity to various sources of deviations has been published (see Andersen, 1973; Andrich & Kline, 1981; Formann, 1981; Gustafsson, 1980b; Hamerle & Tutz, 1980; Martin-Löf, 1973; Molenaar, 1981; Stelzl, 1979; Van den Wollenberg, 1979, 1982; Wright & Stone, 1979). In sum, it can be said that there exist enough methods for diagnosing item bias with the RM.



The recognition of different item parameters in groups of individuals, however, makes it necessary to interpret the causes for observed bias solely on an intuitive basis. If the item material is of such a structure that the LLTM is applicable, the analysis can be done on the level of single cognitive operations, thus revealing the sources of item bias. This was shown in the following paradigmatic study by Piswanger (1975): The set of 42 matrices items analyzed by Formann (1973) was presented to a calibration sample of 2,485 Austrian students aged 14 to 18 and to 159 Togolese and to 200 Nigerian students of the same age group. In each of the three samples the LLTM was applied independently in order to describe item difficulty in terms of the elementary cognitive operations discussed above. The parameter estimates  $\hat{\eta}_j$  obtained are given in Table 1. Since each of the items comprised exactly two components, each item parameter  $b_i$  contained two basic parameters of each of the three factors (*relevant property, rule, direction*). Therefore, one independent normalization condition per factor had to be introduced: The largest  $\eta_j$  within each factor, corresponding to the most difficult operation, was normalized to be zero.

As Table 1 shows, the rank order of the basic parameters was the same in all three samples. With regard to *relevant property* there are no substantial differences. As regards *rule*, the relative difficulty of "superimposition," as compared to "continuation" and "variation," seemed lesser for the Togolese than for the Nigerians and Austrians. This can be understood in view of the fact that operating with number, letter, and symbol sequences has become second nature for individuals in Western civilization (which is true for the Austrians and in part also for the Nigerians, who already have a long-standing European school tradition), whereas "superimposition" is rather an unusual and therefore less salient operation. For the Togolese, whose parents are often still analphabets, there is a less distinct difference in mastery between these operations. The largest discrepancy, however, resulted with respect to *direction*: For the Austrians it was by far easiest to apply a rule in the "horizontal" direction (i.e., always from the left to the right due to the item format). Applying a rule in the "horizontal

Table 1  
Cross-Cultural Comparison of the Basic Parameters  
of the Matrices Test

Factor and Level	Parameter Estimates and 99% Confidence Intervals		
	Austria	Nigeria	Togo
Relevant Property			
Form	$-.77 \pm .04$	$-.94 \pm .15$	$-.82 \pm .16$
Pattern	$-.48 \pm .04$	$-.62 \pm .16$	$-.57 \pm .18$
Number	$-.03 \pm .04$	$-.13 \pm .14$	$-.11 \pm .16$
Array	0	0	0
Rule			
Continuation	$-1.55 \pm .06$	$-1.54 \pm .20$	$-1.14 \pm .22$
Variation	$-1.08 \pm .05$	$-1.11 \pm .17$	$-.66 \pm .19$
Superimposition	0	0	0
Direction			
Horizontal	$-.53 \pm .07$	$-.25 \pm .24$	$-.08 \pm .27$
Horizontal & Vertical	$-.08 \pm .05$	$-.08 \pm .17$	$-.01 \pm .19$
Vertical	0	0	0

and vertical" direction, second in difficulty, was slightly easier than in the "vertical" direction alone, the difference between the two being, however, negligible. For the Nigerians, these differences were much smaller, and for the Togolese, trifling. These findings can also be attributed to the cultural environments: The more a person is rooted in Western civilization with its characteristic reading and writing habits, the more he/she tends to search for lawfulness from the left to the right, relative to other directions. On the other hand, the older generation of Africans, e.g., the parents of the testees, partly still use Arabic writing (from the right to the left). Because of the arrangement of the graphic components, items with a "horizontal" direction can most easily be solved working from the left to the right, and such items are thus harder if the person tries to work in the opposite direction.

Similarly, Whitely and Schneider (1981) have found gender differences with items from the Cognitive Abilities Test (Thorndike & Hagen, 1974); although the overall comparison yielded no significant differences between gender groups, a significant difference was obtained for the operation "spatial distortions." Another recent study describing item bias on the basis of a linear structure of the item parameters is due to Mislevy (1981).

#### Succession in the Rorschach Inkblot Test

In the Rorschach test "succession" means the specific sequence of responses a person gives to one inkblot. Sequential dependencies within the sequence of responses are well known from Rorschach practice (cf. Alcock, 1963, pp. 68-69; Bohm, 1957, pp. 91-92), but the opinions of theoreticians about the Rorschach test differ with respect to the diagnostic significance of succession. Newer psychometrically oriented modifications of the inkblot test avoid this problem by considering only one response per plate (Holtzman, Thorpe, Swartz, & Herron, 1965, pp. 6-7, 11-12). Applications of latent trait theory to the Rorschach test have also been restricted to the first response per inkblot (Fischer, 1968, pp. 118-132; Fischer & Spada, 1970, 1973; Spada & Fischer, 1973). For analyzing complex answer sequences, no appropriate models were available in the past. At first sight, sequential dependencies seem to contradict the assumption of local stochastic independence, which is central to latent trait models. Nevertheless, it will be shown in the following that the LLTM is useful for revealing certain dynamic processes which occur when a person interprets an inkblot.

In spite of the obvious complexity of the phenomenon of succession, there seems to exist some regularity in response sequences, especially regarding "location": "For example, a protocol containing many responses is likely to include a smaller *proportion* of 'whole' responses, since the number of whole responses that can reasonably be perceived in the blots is quite limited, while associations to isolated details of the blots may continue indefinitely" (Anastasi, 1968, p. 497). As soon as a person gives a whole (*W*) response to an inkblot, he/she exhausts her repertory of *W*'s to some extent and will, therefore, in the sequel, tend to give "detail" or "small detail" responses (both denoted here by *D*). This, of course, implies a change in the person (i.e., in his/her behavior), but in a formal sense it can also be postulated that the item difficulty as regards the second response changes, depending on the first response. (For simplicity, considerations are restricted here to two responses per person and plate). The LLTM enables formalization and empirical testing of the following three hypotheses:

- $H_0$ : The first response has no influence on the second response, i.e., the parameters of inkblot difficulty are constant.
- $H_1$ : If person *a* gives a *W* to plate *i*, the difficulty parameter of plate *i* for person *a* increases by a constant amount  $\delta$  as far as the next response is concerned, i.e., the regularity of change can be generalized over individuals and inkblots.

$H_2$ : If person  $a$  gives a  $W$  to plate  $i$ , the difficulty parameter of plate  $i$  for subject  $a$  changes by a constant  $\delta_i$ , i.e., the regularity of change can be generalized across individuals but need not be the same for all inkblots.

Assuming the RM as the response model,  $H_0$  can be formalized by means of the following set of equations, whereby 30 "technical" items ( $T$  items) are assigned to the 20 reactions per person.

$$\begin{aligned}
 b_1 &= b_2 = b_3 = \eta_1 && + c \\
 b_4 &= b_5 = b_6 = \eta_2 && + c \\
 &\vdots && \vdots \\
 b_{28} &= b_{29} = b_{30} = \eta_{10} && + c,
 \end{aligned}
 \tag{15}$$

where  $b_1$  denotes the difficulty of Plate I for the first response,  $b_2$  its difficulty after a previous  $D$ , and  $b_3$  after a  $W$  (i.e., to each condition under which a response can be observed, a  $T$  item parameter is assigned). In the same manner,  $b_4, b_5,$  and  $b_6$  denote the difficulty parameters for the responses to Plate II, and so forth. The basic parameters  $\eta_1, \dots, \eta_{10}$  are measures of the difficulty of each of the 10 inkblots, and  $c$  is the usual normalization constant in the RM. The system of Equations 15 expresses linear constraints on the item parameters  $b_i$  so that the RM becomes an LLTM. In a similar manner,  $H_1$  is formalized by means of the following system of equations:

$$\begin{aligned}
 b_1 &= b_2 = \eta_1 && + c \\
 b_3 &= \eta_1 && + \delta + c \\
 b_4 &= b_5 = \eta_2 && + c \\
 b_6 &= \eta_2 && + \delta + c \\
 &\vdots && \vdots \\
 b_{28} &= b_{29} = \eta_{10} && + c \\
 b_{30} &= \eta_{10} + \delta && + c
 \end{aligned}
 \tag{16}$$

Finally,  $H_2$  can be expressed by

$$\begin{aligned}
 b_1 &= b_2 = \eta_1 && + c \\
 b_3 &= \eta_1 && + \delta_1 + c \\
 b_4 &= b_5 = \eta_2 && + c \\
 b_6 &= \eta_2 && + \delta_2 + c \\
 &\vdots && \vdots \\
 b_{28} &= b_{29} = \eta_{10} && + c \\
 b_{30} &= \eta_{10} + \delta_{10} && + c
 \end{aligned}
 \tag{17}$$

When comparing the systems of Equations 15, 16, and 17, it is seen that they form a hierarchy, with Equations 15 being the strongest and Equations 17 the weakest hypotheses. (Of course, many more hypotheses could also be formulated in the same manner).

These three hypotheses have been tested empirically on the basis of 150 Rorschach protocols (from 50 patients with organic psychosyndrome, 50 neurotics, and 50 healthy control subjects). The protocols comprised one or two responses per person and plate, i.e., the data were incomplete. For estimating the basic parameters  $\eta_i$  and  $\delta$  (or  $\delta_i$ , respectively), the data matrix  $U$  had to be transformed into a new data matrix  $\tilde{U}$  with  $n = 30$   $T$  items (see Table 2). Apart from the observations already missing in  $U$ ,  $\tilde{U}$  was necessarily highly incomplete, since reactions were observed only to either  $T$  items 2 or 3, 5 or 6, and so forth.

Table 2  
Schematic Representation of the Responses of N=150  
Subjects to 10 Rorschach plates (Data Matrix  $U$ ) and Their  
Reinterpretation as Responses to 30  $T$  Items (Data Matrix  $\tilde{U}$ )

Subject	U							$\tilde{U}$									
	Plate							T Item									
	1	1'	2	2'	...	10	10'	1	2	3	4	5	6	...	28	29	30
1	1	0	0	0	...	1	0	1	*	0	0	0	*	...	1	*	0
2	1	1	0	1	...	0	1	1	*	1	0	1	*	...	0	1	*
...																	
149	0	0	1	*	...	0	*	0	0	*	1	*	*	...	0	*	*
150	0	1	0	0	...	1	0	0	1	*	0	0	*	...	1	*	0

\*Denotes missing observation.

Empirical analyses of the data of the 150 individuals under  $H_0$ ,  $H_1$ , and  $H_2$  yielded the estimates and confidence intervals given in Table 3. The plate parameters  $\eta_i$  were normalized such that  $\eta_1 = 0$ . The parameters  $\eta_i$  mirror the fact known from Rorschach practice that  $W$  responses to Plates I and V are quite popular, whereas  $W$  responses to Plates VIII and X are given rather rarely (only by uncritical or by creative persons). The parameter  $\delta$ , estimated under  $H_1$ , proves the truth of the conjecture that a  $W$  as the first response makes the inkblot more difficult with regard to a second  $W$ ; however, this effect is not very strong. The parameters  $\delta_i$  estimated under  $H_2$  reveal a considerable variability of such effects, e.g., for Plate III, second  $W$  responses are very rare, whereas for Plate X there seems to be even a slight (albeit nonsignificant) increase in the tendency to produce  $W$ 's after a first  $W$ .

For an inferential comparison of the three hypotheses, the conditional likelihoods, computed according to Equation 5, were used. These were  $\ln L = -1018.60$  for  $H_0$ ,  $\ln L = -997.58$  for  $H_1$ , and  $\ln L = -971.07$  for  $H_2$ . The asymptotic test statistic for  $H_0$  against  $H_1$ , therefore, was  $\chi^2 = -2(-1018.60 + 997.58) = 42.04$  with  $df = 1$  (because  $H_1$  differs from  $H_0$  by only one additional parameter  $\delta$ ), which was significant at  $\alpha = .01$ . The hypothesis of no effect of the first response on the second one therefore had to be rejected. The test statistic for  $H_1$  against  $H_2$  was  $\chi^2 = -2(-997.58 + 971.07) = 53.02$  with  $df = 9$ , which again was significant at  $\alpha = .01$ . The hypothesis of an effect of the first response which can be generalized over all plates had to be rejected. This is not surprising in view of the complexity of the process of interpreting an inkblot.

Table 3  
 Rorschach Succession: Basic Parameter Estimates  
 Under  $H_0$ ,  $H_1$ , and  $H_2$

Parameter	Parameter Estimates and 99% Confidence Intervals		
	$H_0$	$H_1$	$H_2$
Plate			
Parameter			
1	0	0	0
2	2.03 ± .59	2.24 ± .60	2.06 ± .77
3	.94 ± .60	.99 ± .61	-.10 ± .92
4	.44 ± .63	.52 ± .64	.35 ± .85
5	-.11 ± .70	-.04 ± .71	-.67 ± 1.03
6	.96 ± .61	1.11 ± .62	.56 ± .82
7	2.47 ± .60	2.75 ± .62	2.65 ± .77
8	3.54 ± .63	3.85 ± .65	3.64 ± .79
9	2.02 ± .60	2.26 ± .61	2.10 ± .77
10	2.90 ± .60	3.17 ± .62	3.11 ± .77
Transfer			
Parameter			
1	--	.74 ± .30	.23 ± .94
2	--	--	.46 ± .83
3	--	--	2.37 ± .87
4	--	--	.55 ± .90
5	--	--	1.50 ± 1.10
6	--	--	1.56 ± .87
7	--	--	-.12 ± .99
8	--	--	.39 ± 1.09
9	--	--	.28 ± .94
10	--	--	-.23 ± .90

Measurement of Change

It is a characteristic of a successful empirical science that it allows valid predictions of reality and, thereby, the purposeful intervention of people on their living conditions. Applied to psychology this means that, e.g., the validity of predictions about the effects of educational or clinical treatments constitutes an important criterion for the scientific character of applied psychology. As will be shown in the following, latent trait theory provides a solid basis for attaining this goal.

Linear Logistic Models with Relaxed Assumptions

Parallel to the development of the LLTM described here, models for measuring attitudinal and behavioral changes due to the influence of mass communications have been developed; these models

can also be used for quantifying other kinds of treatment effects (Fischer, 1972, 1976, 1977a, 1977b, 1980). Formally, they can be considered LLTMs for incomplete data.

Suppose that the reactions of  $N$  individuals to  $n$  items (or more generally:  $n$  criteria, such as the presence or absence of certain clinical symptoms) have been observed at several points of time  $t_1, t_2, \dots, t_h$  with the aim of studying the effects of given treatments or treatment combinations. Let the respective data matrices be denoted by  $U_x$ , with elements  $u_{iax}$ ,  $x = 1, \dots, h$ . The probability of reaction “+” of person  $a$  to item (or criterion)  $i$  at time  $t_x$  will be described by the following ICC:

$$P(+ | i, a, t_x) = \frac{\exp[\theta_{ia} + (t_x - t_1)\delta_a]}{1 + \exp[\theta_{ia} + (t_x - t_1)\delta_a]}, \quad [18]$$

$$\delta_a = \sum_j q_{aj} \eta_j + \tau \quad [19]$$

for  $a = 1, \dots, N, i = 1, \dots, n, x = 1, \dots, h$  and  $j = 1, \dots, m$ , where

$\theta_{ia}$  is the latent ability or tendency of person  $a$  to give the reaction “+” to item (or criterion)  $i$  at time  $t_1$ ;

$q_{aj}$  is the dose of treatment  $j$  as applied to person  $a$ ;

$\eta_j$  is the effect of treatment  $j$ ;

$\tau$  is the “trend,” i.e., the sum of effects which are independent of the treatments; and

$\delta_a$  is the total amount of change in person  $a$  caused by treatments and trend.

This ICC differs from the customary ICCs because the individuals are not characterized with respect to just one latent dimension but to a multidimensional latent space. The items, therefore, need not be unidimensional, i.e., conforming to the RM or to some other well-known latent trait model, but may be heterogeneous. Hence, it may hold that, for example, one person  $a$  should have a pronounced tendency to show symptom  $i$  but a small likelihood for showing symptom  $l$ , while the reverse is true for another person  $a'$ . The model defined by Equations 18 and 19 has been termed the “linear logistic model with relaxed assumptions” (LLRA) because it drops the restrictive assumption of unidimensionality, which is difficult to attain in many educational or clinical settings. Only the treatment effects are measured by scalar parameters in a single dimension, and at the same time also the amount of change in any person  $a$  as the sum of treatment effects and trend.

The philosophy underlying this model is that in applied science simple answers to simple questions, as, e.g., “What is the relative effect of treatment  $j$  in comparison to treatment  $l$ ?” are often needed. In order to allow, at least in principle, an answer to such questions, an additive combination of the effect parameters  $\eta_j$ , weighted by the dose measures, has been assumed in Equation 19. The additivity implies a “concatenation operation” for the treatments and entails ratio-scale properties for the  $\eta_j$ 's and  $\tau$ . These are very strong assumptions, of course, but they may serve as an  $H_0$  against other more complicated assumptions; so, for instance, interaction terms might be included in Equation 19, or nonlinear dose response curves. Within the scope of the present paper, however, it is not possible to discuss the complications of the model ensuing from such generalizations.

It is easily seen that the model defined by Equations 18 and 19 can be reduced to an LLTM: Because of the local stochastic independence and the multidimensional parameterization, the reactions  $u_{iax}$  and  $u_{jax}$  of the same person  $a$  are independent of each other as if they stemmed from two different individuals. All the elements of one data matrix  $U_x$  can therefore be considered, in a purely formal sense, as the responses of  $nN$  “technical” persons ( $T$  persons) with person parameters  $\theta_g^*$ ,  $g = (a - 1)n + i$ , i.e., with  $g$  running from 1 through  $nN$ , given to  $h$  items with the parameters  $(t_x - t_1)\delta_a$ ,

$$P(+|i, a, t_x) = P(+|g, t_x) = \frac{\exp[\theta_g^* + (t_x - t_1)\delta_a]}{1 + \exp[\theta_g^* + (t_x - t_1)\delta_a]} \quad [20]$$

This transforms the ICCs of Equation 18 into those of an LLTM with incomplete data. The weights in the linear Equation 19 seem to depend on the individual  $a$  (i.e., on  $\theta_a$ ) but, as a matter of fact, only depend on the treatment doses prespecified in the design. Hence, it is in principle possible to estimate the parameters  $\eta_j$  and  $\tau$  and to test hypotheses by the same methods as in the LLTM described above. It can be shown that the estimation equations become much simpler if the design comprises only two points of time, i.e., if there is a pretest-posttest design (Fischer, 1977a, 1977b, 1980, 1982). Mainly for this reason almost all the applications thus far have been restricted to two times of observation.

### Applications of the Model

A detailed exposition of the LLRA, of its properties, and of the many applications is not possible here due to lack of space (some of them are mentioned in Fischer, 1978); in order to illustrate the potential uses of the LLRA, only two cases will be described. Rop (1977) studied the effects of three cognitive preschool education programs on the development of intelligence in Viennese kindergarten children. A primary aim of the study was to find the maximal range of generalizability of the treatment effects across different subsamples of children and of items (i.e., latent ability dimensions). The effect parameters  $\eta_j$  and  $\tau$  could, in fact, be generalized over all the samples of children, but the effects of the programs turned out to be significantly different in three areas of cognitive abilities (verbal intelligence, vocabulary, and nonverbal intelligence). One notable result was that the early reading program was by far the least effective and had a significant impact only on the domain of vocabulary items. The two other programs (logical thinking and verbal training) proved effective in all three ability domains, i.e., also with respect to nonverbal intelligence.

Heckl (1976) investigated the effects of speech therapies in children with speech handicaps, using a very heterogeneous battery of test items, symptoms, and anamnestic data as "criteria" of change. No difference was found between effects of the three forms of therapy used, but there were different effects in children with severe speech disorders and in those with only slight handicaps. The trend parameter  $\tau$ , for instance, differed nonsignificantly from zero for children with severe speech disorders, which implies that these children made no significant progress (within the time period studied and with respect to the criteria used), unless they obtained therapy. Among the slightly handicapped, on the other hand, there was a significant positive trend independent of the treatments. Heckl also tested the linearity of the dose response curve, a hypothesis which had to be rejected: If therapy was applied too frequently (i.e., twice a week), then its effect was less than that of two equivalent therapy sessions in subsequent weeks, in other words, a satiation phenomenon was observed.

These remarks on applications of linearly restricted logistic models to the measurement of change should show that latent trait theory can go far beyond mere quantification of item and person parameters and can be used as a means of hypothesis testing. Further details can be found in the literature mentioned. Similar developments have also recently evolved from the theory of log-linear models (see Breslow, 1976; Koch, Landis, Freeman, Freeman, Jr., & Lehnen, 1977; Marascuilo & Serlin, 1979; Plewis, 1981).

### Discussion

What conclusions can be drawn about the usefulness of linearly restricted logistic models (LLTM and LLRA)? The hopes originally harbored—that the sources of the cognitive complexity of certain

item materials could largely be clarified—have not been entirely fulfilled, inasmuch as the LR tests of the LLTM against the RM (as the alternative hypothesis) mostly turned out to be significant, so that the LLTM had to be rejected. On the other hand, such statistical significances ought not to be over-rated, because in many cases relatively large samples of data were used for testing hypotheses about only a few parameters, i.e., the tests were rather powerful; moreover, ultimately any significance criterion is arbitrary. In order to assess the psychological relevance of the applications, it therefore seems to be decisive whether the approximate constancy of the basic parameters across person and item samples suffices for practical test construction purposes and/or from the point of view of psychological theorizing.

An interesting contribution to this question stems from Nährer (1977, 1980). He constructed a few new items, each for several of the item materials described above, and predicted the item parameters using the original estimates of the basic parameters. He then carried out a cross-validation by comparing the predicted parameters with the estimates  $\hat{b}_i^n$  obtained in new samples of persons. This approach is remarkable because it amounts to real predictions—new items presented to new samples of individuals—which are otherwise unusual in psychological research. The approximate correspondence he was able to observe does not express model fit in the traditional sense, but is rather the result of the validity of the underlying structural hypotheses which, of course, can never be perfect. Even in cases of an unsatisfactory conformity of the model to the data, the mere formulation of those hypotheses which are needed for working with the LLTM leads to a clearer understanding of the substantive problems, as was illustrated in the section on item bias.

However, researchers should not be content with a rough fit of the model and with considering the correlation between parameter estimates  $\hat{b}_i$  based on Equation 4 and estimates  $\hat{b}_i^n$  within the same sample as an indicator of a sufficient explanatory value of the model (as was done in some early applications of the LLTM). Such correlations will practically always occur, their sizes depending essentially on the range of the difficulty parameters of the items. Finally, it is meaningless to test such a correlation for significance since the  $H_0$  that the LLTM holds cannot be expressed by “ $r = 0$ ,” but would have to be “ $r = 1$ ,” for which no adequate test statistics exist.

Another temptation that should be resisted is to formulate too large a hierarchy of structural hypotheses, from very simple ones up to the “saturated” hypothesis with a maximum number of identifiable parameters, and to make a large number of corresponding statistical tests on the same set of data. This would favor the occurrence of random results not supported by prior psychological hypotheses.

Compared with the LLTM, the LLRA at the present state of knowledge can be rated more optimistically. It has often been possible in empirical applications to retain null hypotheses of the generalizability of results over item or person samples. This may be a result of the high flexibility of the model due to its multidimensional parameterization. At the same time, LLRA applications are not limited to homogeneous item materials (as is the case with the LLTM), but range from intelligence testing over attitude questionnaires to clinical symptoms or behavioral criteria. The focus of the applications of linearly constrained logistic models therefore is shifting more and more to the problem area of measuring change.

#### References

- Alcock, T. *The Rorschach in practice*. London: Tavistock, 1963.
- Anastasi, A. *Psychological testing* (3rd ed.). New York: MacMillan, 1968.



- Andersen, E. B. A goodness of fit test for the Rasch model. *Psychometrika*, 1973, 38, 123-140.
- Andersen, E. B. *Discrete statistical models with social science applications*. Amsterdam: North Holland Publishing Company, 1980.
- Andrich, D., & Kline, P. Within and among population fit with the simple logistic model. *Educational and Psychological Measurement*, 1981, 41, 35-48.
- Bock, R. D., & Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 1981, 46, 443-459.
- Bohm, E. *Lehrbuch der Rorschach-Psychodiagnostik*. Bern: Huber, 1957.
- Breslow, N. Regression analysis of the log odds ratio: A method for retrospective studies. *Biometrics*, 1976, 32, 409-416.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, 39, 1-38.
- Falmagne, J. C. Unpublished manuscript cited in H.-Ch. Micko, Eine Verallgemeinerung des Messmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. *Psychologische Beiträge*, 1970, 12, 4-22.
- Fischer, G. H. Neue Entwicklungen in der psychologischen Testtheorie. In G. H. Fischer (Ed.), *Psychologische Testtheorie*. Bern: Huber, 1968.
- Fischer, G. H. A measurement model for the effect of mass-media. *Acta Psychologica*, 1972, 36, 207-220.
- Fischer, G. H. The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 1973, 36, 359-374.
- Fischer, G. H. *Einführung in die Theorie psychologischer Tests*. [Introduction to the theory of psychological tests.] Bern: Huber, 1974.
- Fischer, G. H. Some probabilistic models for measuring change. In D. N. M. de Gruijter & L. J. Th. van der Kamp (Eds.), *Advances in psychological and educational measurement*. New York: Wiley, 1976.
- Fischer, G. H. Linear logistic models for the description of attitudinal and behavioral changes under the influence of mass communication. In W. F. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology*. Bern: Huber, 1974/New York: Wiley, 1977. (a)
- Fischer, G. H. Linear logistic test models: Theory and application. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning*. Bern: Huber, 1977. (b)
- Fischer, G. H. Probabilistic test models and their application. *The German Journal of Psychology*, 1978, 2, 298-319.
- Fischer, G. H. Some latent trait models for measuring change in qualitative observations. In D. J. Weiss (Ed.), *Proceedings of the 1979 Computerized Adaptive Testing Conference*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory, 1980.
- Fischer, G. H. On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, 1981, 46, 59-77.
- Fischer, G. H. *Logistic latent trait models with linear constraints: Formal results and typical applications* (Research Bulletin No. 24). Vienna: University of Vienna, Institute of Psychology, 1982. (*Psychometrika*, in press)
- Fischer, G. H., & Formann, A. K. *An algorithm and a FORTRAN program for estimating the item parameters of the linear logistic test model* (Research Bulletin No. 11). Vienna: University of Vienna, Institute of Psychology, 1972.
- Fischer, G. H., & Pendl, P. Individualized testing on the basis of the dichotomous Rasch model. In L. J. Th. van der Kamp, W. F. Langerak, & D. N. M. de Gruijter (Eds.), *Psychometrics for educational debates*. New York: Wiley, 1980.
- Fischer, G. H., & Scheiblechner, H. Algorithmen und Programme für das probabilistische Testmodell von Rasch. [Algorithms and programs for the probabilistic test model of Rasch.] *Psychologische Beiträge*, 1970, 12, 23-51.
- Fischer, G. H., & Spada, H. Die metrische Analyse eines projektiven Formdeutetests. *Psychologische Beiträge*, 1970, 12, 65-82.
- Fischer, G. H., & Spada, H. *Die psychometrischen Grundlagen des Rorschachtests und der Holtzman Inkblot Technique*. Bern: Huber, 1973.
- Formann, A. *Die Konstruktion eines neuen Matrixentests und die Untersuchung des Lösungsverhaltens mit Hilfe des linearen logistischen Testmodells*. Unpublished doctoral dissertation, University of Vienna, 1973.
- Formann, A. K. Über die Verwendung von Items als Teilungskriterium für Modellkontrollen im Modell von Rasch. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1981, 28, 541-560.
- Formann, A. K., & Piswanger, K. *Wiener Matrizen-Test. Ein Rasch-skaliertes sprachfreier Intelligenztest*. Weinheim: Beltz Test, 1979.
- Gustafsson, J.-E. A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement*, 1980, 40, 377-385. (a)

- Gustafsson, J.-E. Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 1980, 33, 205-233. (b)
- Haberman, S. J. Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, 1977, 5, 815-841.
- Habon, M. *Sprachfreie Tests in der pädagogischen Diagnostik. Regelgeleitete Aufgabenkonstruktion durch formale Modelle*. Unpublished master's thesis, Philipps-University of Marburg, 1981.
- Hamerle, A., & Tutz, G. Goodness of fit tests for probabilistic measurement models. *Journal of Mathematical Psychology*, 1980, 21, 153-167.
- Haudek, U. *Ein Modell der Schwierigkeit von Denkaufgaben*. Unpublished doctoral dissertation, University of Vienna, 1973.
- Heckl, U. *Therapieerfolge bei der Behandlung sprachgestörter Kinder*. Unpublished doctoral dissertation, University of Vienna, 1976.
- Heinrich, I. Beeinflussung des logischen Schlussprozesses durch semantische Variation seiner Elemente. *Psychologische Beiträge*, 1975, 17, 497-518.
- Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Heron, E. W. *Inkblot perception and personality* (2nd ed.) Austin: University of Texas Press, 1965.
- Horn, W. *Leistungsprüfsystem (LPS)*. Göttingen: Hogrefe, 1962.
- Kearney, G. E. *The cognitive ability of aboriginal Australians*. Unpublished doctoral dissertation, University of Queensland, 1966.
- Kempf, W. F. Probabilistische Modelle experimentalphysikalischer Versuchssituationen. *Psychologische Beiträge*, 1972, 14, 16-37.
- Kempf, W. F. Dynamic models for the measurement of 'traits' in social behavior. In W. F. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology*. Bern: Huber, 1974/New York: Wiley, 1977.
- Koch, G. G., Landis, R. J., Freeman, J. L., Freeman, D. H. Jr., & Lehnen, R. G. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 1977, 33, 133-138.
- Kubinger, K. D. Das Problemlöseverhalten bei der statistischen Auswertung psychologischer Experimente. Ein Beispiel hochschuldidaktischer Forschung. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1979, 26, 467-495.
- Kubinger, K. D. Die Bestimmung der Effektivität universitärer Lehre unter Verwendung des linearen logistischen Testmodells von Fischer—Neue Ergebnisse. [The application of Fischer's linear logistic test model in order to evaluate the effectiveness of higher education courses—Some new results.] *Archiv für Psychologie*, 1980, 133, 69-79.
- Lord, F. M. *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum, 1980.
- Marascuilo, L. A., & Serlin, R. C. Tests and contrasts for comparing change parameters for a multiple sample McNemar data model. *The British Journal of Mathematical and Statistical Psychology*, 1979, 32, 105-112.
- Martin-Löf, P. *Statistiska modeller. Anteckningar från seminarier läsåret 1969-70 utarbetade av Rolf Sundberg, 2:a uppl. [Statistical models. Notes from seminars 1969-70 by Rolf Sundberg, 2nd ed.]* Stockholm: University of Stockholm, 1973.
- Mellenbergh, G. J. Applicability of the Rasch model in two cultures. In L. J. Cronbach & P. J. D. Drenth (Eds.), *Mental tests and cultural adaptation*. The Hague: Mouton, 1972.
- Micko, H.-Ch. A psychological scale for reaction time measurement. *Acta Psychologica*, 1969, 30, 324-335.
- Micko, H.-Ch. Eine Verallgemeinerung des Messmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. *Psychologische Beiträge*, 1970, 12, 4-22.
- Mislevy, R. J. A general linear model for the analysis of Rasch item threshold estimates. Unpublished doctoral dissertation, University of Chicago, 1981.
- Molenaar, I. *Some improved diagnostics for failure of the Rasch model*. Groningen: Heymans Bulletin, Rijksuniversiteit Groningen, Psychologische Instituten, 1981.
- Nährer, W. *Modellkontrollen bei der Anwendung des linearen logistischen Modells in der Psychologie*. Unpublished doctoral dissertation, University of Vienna, 1977.
- Nährer, W. Zur Analyse von Matrizenaufgaben mit dem linearen logistischen Testmodell. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1980, 27, 553-564.
- Nenty, H. J., & Dinero, T. E. A cross-cultural analysis of the fairness of the Cattell culture fair intelligence test using the Rasch model. *Applied Psychological Measurement*, 1981, 5, 355-368.
- Piswanger, K. *Interkulturelle Vergleiche mit dem Matrizen-test von Formann*. Unpublished doctoral dissertation, University of Vienna, 1975.
- Plewis, I. A comparison of the approaches to the analysis of longitudinal categorical data. *The British Journal of Mathematical and Statistical Psychology*, 1981, 34, 118-123.
- Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Paedagogiske Institut, 1960.
- Rasch, G. An informal report on a theory of objectivity in comparisons. In L. J. Th. van der Kamp & C. A. J. Vlek (Eds.), *Measurement theory*. Leiden:

- University of Leiden, Psychological Institute, 1967. (Proceedings of the NUFFIC international summer session in science at "Het Oude Hof," The Hague, July 1966.)
- Rasch, G. *A mathematical theory of objectivity and its consequences for model construction*. Paper presented at the European Meeting on Statistics, Econometrics, and Management Science, Amsterdam, The Netherlands, September 1968.
- Rasch, G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In M. Blegvad (Ed.), *The Danish yearbook of philosophy*. Copenhagen: Munksgaard, 1977.
- Rop, I. The application of a linear logistic model describing the effects of pre-school curricula on cognitive growth. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning*. Bern: Huber, 1977.
- Scandura, J. M. *Structural learning I. Theory and research*. New York: Gordon & Breach, 1973.
- Scheiblechner, H. *CML-parameter-estimation in a generalized multifactorial version of Rasch's probabilistic measurement model with two categories of answers* (Research Bulletin No. 4). Vienna: University of Vienna, Institute of Psychology, 1971.
- Scheiblechner, H. Das Lernen und Lösen komplexer Denkaufgaben. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1972, 19, 476-506.
- Scheuneman, J. *A new method of assessing bias in test items*. Paper presented at the meeting of the American Educational Research Association, Washington, April 1975. (ERIC Document Reproduction Service No. ED 106-359)
- Scheuneman, J. Latent trait theory and item bias. In L. J. Th. van der Kamp, W. F. Langerak, & D. N. M. de Gruijter (Eds.), *Psychometrics for educational debates*. New York: Wiley, 1980.
- Spada, H. *Grundlagenforschung zur Intelligenzdiagnostik*. Unpublished doctoral dissertation, University of Vienna, 1969.
- Spada, H. Intelligenztheorie und Intelligenzmessung. *Psychologische Beiträge*, 1970, 12, 83-96.
- Spada, H. *Modelle des Denkens und Lernens*. Bern: Huber, 1976.
- Spada, H. Logistic models of learning and thought. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning*. Bern: Huber, 1977.
- Spada, H., & Fischer, G. H. Latent trait models and the problem of measurement in projective techniques (Rorschach, Holtzman). In A. Serrate (Ed.), *Rorschachiana X. Beiheft zur Schweizerischen Zeitschrift für Psychologie und ihre Anwendungen*, No. 55. Bern: Huber, 1973, (Paper presented at the VIIIth International Congress of Rorschach and Projective Methods, Zaragoza, Spain, 1971.)
- Spada, H., Fischer, G. H., & Heyner, W. Denkopoperationen und Lernprozesse bei der Lösung von Problemstellungen aus der Mechanik. In L. H. Eckensberger & V. H. Eckensberger (Eds.), *Bericht über den 28. Kongress der Deutschen Gesellschaft für Psychologie, Saarbrücken 1972*. Göttingen: Hogrefe, 1974.
- Stelzl, I. Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothesen zu prüfen? Ein Bericht über Simulationsstudien mit inhomogenen Daten. [Is the Rasch model sensitive to tests of homogeneity?] *Zeitschrift für Experimentelle und Angewandte Psychologie*, 1979, 26, 652-672.
- Suppes, P. Learning theory for probabilistic automata and register machines with applications to educational research. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning*. Bern: Huber/Fort Lee NJ: Update Publishing International, 1977.
- Thissen, D. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 1982, 175-186.
- Thorndike, R. L., & Hagen, E. *Cognitive abilities test*. New York: Houghton-Mifflin, 1974.
- Van den Wollenberg, A. *The Rasch model and time-limit tests*. Nijmegen: Stichting Studentenpers Nijmegen, 1979.
- Van den Wollenberg, A. Two new test statistics for the Rasch model. *Psychometrika*, 1982, 47, 123-140.
- Wainer, H., Morgan, A., & Gustafsson, J.-E. A review of estimation procedures for the Rasch model with an eye towards longish tests. *Journal of Educational Statistics*, 1980, 5, 35-64.
- Whitely, S. E., & Schneider, L. M. Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement*, 1981, 5, 383-397.
- Wright, B. D., Mead, R., & Draba, R. *Detecting and correcting test item bias with a logistic response model* (Research Memorandum No. 22). Chicago: The University of Chicago, Department of Education, Statistical Laboratory, 1976.
- Wright, B. D., & Panchapakesan, N. A procedure for sample-free item analysis. *Educational Psychological Measurement*, 1969, 2, 23-48.
- Wright, B. D., & Stone, M. H. *Best test design*. Chicago: Mesa Press, 1979.

**Authors' Address**

Send requests for reprints or further information to Gerhard H. Fischer or Anton K. Formann, Institute of Psychology, University of Vienna, Liebiggasse 5, A-1010 Wien, Austria.

**Acknowledgments**

*The present research was supported by the Österreichische Forschungsgemeinschaft under Grant No. 01/0054.*