

Measuring Change By Means of a Hybrid Variant of the Linear Logistic Model With Relaxed Assumptions

Anton K. Formann and Christiane Spiel
University of Vienna

The linear logistic model with relaxed assumptions (LLRA) was developed for measuring changes in qualitative data. It assumes item-specific person parameters and thus does not require homogeneous items to be presented to the persons at two points in time. The hybrid variant of this model maintains the multidimensionality of the person parameters, but it allows for different sets of items each of which is presented only once. In the model, a Rasch homogeneous item at t_2 with possibly differing difficulty corresponds to each

item at t_1 . A short description of both models is followed by a first application of the hybrid LLRA to empirical data from a study on text comprehension. This example not only serves to demonstrate possible results when applying the LLRA, but is also used to outline the principle of hypothesis testing and model controls. *Index terms: dichotomous data, linear logistic model, measuring change, Rasch model, text comprehension.*

The quantification of change has become a topic of great interest in many fields. Traditionally, change is defined in terms of differences of values measured at different points in time. Applying this concept to the social sciences (see Harris, 1963) would require a natural metric of those persons' scores which are used to infer the changes. However, most of the observations collected in the social sciences are qualitative, and thus lack a natural metric. On the other hand, quantitative statements about change are desired which, by the traditional method, cannot be obtained other than by an artificial quantification of the observations.

To overcome these problems, special models can be formulated for measuring change in qualitative data. One of these models, the "linear logistic model with relaxed assumptions" (LLRA; Fischer, 1972, 1976; for similar models see Breslow, 1976; Duncan, 1985a, 1985b; Koch, Landis, Freeman, Freeman, & Lehnen, 1977; Marascuilo & Serlin, 1979; Plewis, 1981), is based on Rasch's (1960/1980) concept of specific objectivity which is realized in his well-known one-parameter logistic model for item analysis. In contrast to the latter model, the LLRA does not require homogeneous categorical items, that is, items which measure one and the same latent trait.

The LLRA

In the simplest case, there are two groups of persons—the experimental group and the control group. The reactions of all persons to the same k dichotomous items (e.g., presence vs. absence of a certain

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 13, No. 1, March 1989, pp. 91–103
© Copyright 1989 Applied Psychological Measurement Inc.
0146-6216/89/010091-13\$1.90

clinical symptom) are denoted I_i ($i = 1, \dots, k$) and are observed at two points in time, t_1 and t_2 . Whereas all members of the control group P_v ($v = 1, \dots, n$) remain untreated, all members of the experimental group P_v ($v = n + 1, \dots, N$) receive a certain treatment (e.g., therapy) between t_1 and t_2 . Assuming logistic item characteristic curves (ICCs) for all items I_i ($i = 1, \dots, k$), the following equations describe both groups with respect to their probabilities of positive reactions (“+”) at t_1 , p_{vi1} , and at t_2 , p_{vi2} :

$$\begin{array}{l} \text{control group, } t_1 \\ (v = 1, \dots, n) \end{array} \quad p_{vi1} = \frac{\exp(\xi_{vi})}{1 + \exp(\xi_{vi})} \quad , \quad (1)$$

$$\begin{array}{l} \text{control group, } t_2 \\ (v = 1, \dots, n) \end{array} \quad p_{vi2} = \frac{\exp(\xi_{vi} + \tau)}{1 + \exp(\xi_{vi} + \tau)} \quad , \quad (2)$$

$$\begin{array}{l} \text{experimental group, } t_1 \\ (v = n + 1, \dots, N) \end{array} \quad p_{vi1} = \frac{\exp(\xi_{vi})}{1 + \exp(\xi_{vi})} \quad , \quad (3)$$

$$\begin{array}{l} \text{experimental group, } t_2 \\ (v = n + 1, \dots, N) \end{array} \quad p_{vi2} = \frac{\exp(\xi_{vi} + \eta + \tau)}{1 + \exp(\xi_{vi} + \eta + \tau)} \quad , \quad (4)$$

where ξ_{vi} is the latent tendency of a positive reaction to item I_i at t_1 of person P_v ,

η is the effect of the treatment given to the experimental group, and

τ is the effect of those uncontrollable influences (“trend”) not attributable to the treatment which are effective in the control as well as in the experimental group.

Persons are characterized not by a one-dimensional parameter but by a k -dimensional one, $\xi_{v1}, \dots, \xi_{vk}$. This means that each item I_i may—but need not—measure its specific latent dimension, ξ_i , where each person P_v has his or her own position, ξ_{vi} . As a consequence, the items may be heterogeneous, which is in contrast to usual item response theory assumptions, and justifies the name of the model. Clearly, the possibility of using nonhomogeneous items is very advantageous: It is no longer necessary to resort to known homogeneous test materials or to develop with great effort new ones; each collection of items (symptoms, verbal reactions, questionnaire data, etc.) may become the basis for quantifying changes. Thus, the LLRA is rather flexible with respect to its application, posing only modest demands concerning the data.

The treatment effect and the trend are assumed to be independent of the items; that is, the items have identical sensitivity with respect to the changes. (This assumption is hypothetical; it can and must be tested in each empirical application of the LLRA.) An additive concatenation is assumed for the parameters, leading to a ratio scale for η and τ whose scaling constant depends on the time interval ($t_2 - t_1$) and the units of the treatment.

The structural parameters, η and τ , may be separated from the incidental person parameters, ξ , a fact which can be used to estimate η and τ independently of the multidimensional person parameters and their distribution according to the conditional maximum likelihood (CML) method. In principle, the separability of the parameters can be seen from the log odds ratios

$$\ell_v = \log \left[\frac{p_{vi2}(1 - p_{vi1})}{(1 - p_{vi2})p_{vi1}} \right] = \tau \quad (v = 1, \dots, n) \quad (5)$$

and

$$\ell_v = \log \left[\frac{p_{vi2}(1 - p_{vi1})}{(1 - p_{vi2})p_{vi1}} \right] = \eta + \tau \quad (v = n + 1, \dots, N) \quad . \quad (6)$$

Finally, it should be noted that a variety of hypotheses for η and τ can be tested on the basis of the likelihood ratio statistic, for example:

1. The effect of the treatment and/or the trend is 0, that is, $\eta = 0$ and/or $\tau = 0$.
2. The items have different sensitivity so that the effect of the treatment and the trend is not independent of the items, that is, $\eta(\text{item } I_i) \neq \eta(\text{item } I_j)$ and/or $\tau(\text{item } I_i) \neq \tau(\text{item } I_j)$.
3. The effect of the treatment and the trend is not independent of the persons (e.g., females and males react differently to one and the same therapy), that is, $\eta(\text{group } 1) \neq \eta(\text{group } 2)$ and/or $\tau(\text{group } 1) \neq \tau(\text{group } 2)$.

The simple version of the LLRA described by Equations 1 through 4 and providing for only two groups of persons can easily be generalized to more than two groups with differing treatments, or even to treatment combinations of varying levels. To cover all possible cases, it is advantageous to write the model equations for person P_v as follows:

$$t_1 \quad p_{vi1} = \frac{\exp(\xi_{vi})}{1 + \exp(\xi_{vi})} \quad , \quad (7)$$

and

$$t_2 \quad p_{vi2} = \frac{\exp(\xi_{vi} + \delta_v)}{1 + \exp(\xi_{vi} + \delta_v)} \quad , \quad (8)$$

where ξ_{vi} is defined as above and δ_v is the latent change for person P_v from t_1 to t_2 ,

$$\delta_v = \sum_{j=1}^m q_{vj} \eta_j + \tau \quad , \quad (9)$$

where η_j is the (unknown) effect of treatment T_j ,

q_{vj} is the (known) amount of this treatment, given to person P_v , and

τ is the trend.

To incorporate the pairwise interaction effects γ_{jh} of two treatments T_j and T_h if applied simultaneously, Equation 9 can be extended to

$$\delta_v = \sum_{j=1}^m q_{vj} \eta_j + \sum_{j=1}^{m-1} \sum_{h=j+1}^m q_{vj} q_{vh} \gamma_{jh} + \tau \quad . \quad (10)$$

Unfortunately, the application of the LLRA may incur several problems. First, the assumption may be violated that the items—being presented twice—can be considered to have been answered independently at t_2 , because this would exclude potential effects of learning and memory. Second, the treatment's effects and the trend may not be extremely large, because then the corresponding parameter estimates tend to infinity (as do their standard deviations, which can be determined using the information matrix). This is because the parameter estimation according to the CML principle only uses that part of the data per item for which changes were observable (at t_1 “+” and at t_2 “-”, or at t_1 “-” and at t_2 “+”), whereas the remaining data showing no changes are ignored. Thus, when there are changes solely in one direction, the parameter estimates diverge, so that the quantification of the treatments' effects and the trend becomes impossible. (For details, see the Appendix.) Besides, the CML approach has a further disadvantageous byproduct: If each item shows no or merely minor changes from t_1 to t_2 , the parameter estimates depend only on a very small portion of the data, that is, the data become relatively uninformative; as a consequence, the parameter estimates may become extremely inaccurate (which can be seen from their variances) even if a large sample of persons was available.

The Hybrid Variant of the LLRA

It can be concluded that in its original formulation, the LLRA is of limited usefulness. However, there exists a further variant of this model tailored to that special case where changes are expected to be

very massive and/or where presentation of the same items twice seems to be problematic; it may also be helpful in producing greater (virtual) effects in the manifest data, so that even small latent changes can be estimated with greater accuracy.

This model was named the “hybrid LLRA” (see Fischer, 1977, pp. 221–222) and combines the idea of multidimensional person parameters from the LLRA with that of pairs of items having differing difficulty and being homogeneous in the sense of Rasch’s (1960/1980) logistic model for item analysis: If massive effects can be presumed, the more difficult item out of each pair of homogeneous items is presented to the persons at that point in time where the probability of positive reactions is expected to be higher, so that only a certain portion of the reactions will be positive, but not all; thus, ceiling effects are precluded for the items, and the changes from t_1 to t_2 are compensated at least partially by the difference of the difficulties of both items. If small effects are expected such that the data would become rather uninformative under the conventional LLRA paradigm because of no changes, or if the presentation of the same items twice does not seem to be indicated per se, for about half of the pairs of “parallel” items the easier, and for about half of them the more difficult item is presented at both points in time; this artificially increases the number of observable changes, and the estimates of the latent effect parameters become more precise.

For the most simple case (i.e., for two groups of persons), the control group containing the persons P_v ($v = 1, \dots, n$) and the experimental group consisting of the persons P_v ($v = n + 1, \dots, N$) fully parameterized, the specification equations of the hybrid variant of the LLRA are as follows:

$$\begin{array}{l} \text{control group, } t_1 \\ (v = 1, \dots, n) \end{array} \quad p_{vi1} = \frac{\exp(\xi_{vi} - \sigma_i)}{1 + \exp(\xi_{vi} - \sigma_i)} \quad (11)$$

$$\begin{array}{l} \text{control group, } t_2 \\ (v = 1, \dots, n) \end{array} \quad p_{vi2} = \frac{\exp(\xi_{vi} + \tau - \sigma'_i)}{1 + \exp(\xi_{vi} + \tau - \sigma'_i)} \quad (12)$$

$$\begin{array}{l} \text{experimental group, } t_1 \\ (v = n + 1, \dots, N) \end{array} \quad p_{vi1} = \frac{\exp(\xi_{vi} - \sigma_i)}{1 + \exp(\xi_{vi} - \sigma_i)} \quad (13)$$

$$\begin{array}{l} \text{experimental group, } t_2 \\ (v = n + 1, \dots, N) \end{array} \quad p_{vi2} = \frac{\exp(\xi_{vi} + \eta + \tau - \sigma'_i)}{1 + \exp(\xi_{vi} + \eta + \tau - \sigma'_i)} \quad (14)$$

where ξ_{vi} , τ , and η are defined as above, σ_i is the difficulty of item I_i presented at t_1 to the persons, and σ'_i is the difficulty of the “parallel” item I'_i presented at t_2 and measuring the same latent dimension as does item I_i .

Calculating the log odds ratios by analogy to Equations 5 and 6,

$$\ell_{vi} = \log \left[\frac{p_{vi2}(1 - p_{vi1})}{(1 - p_{vi2})p_{vi1}} \right] = \tau - (\sigma'_i - \sigma_i) \quad (v = 1, \dots, n) \quad (15)$$

and

$$\ell_{vi} = \log \left[\frac{p_{vi2}(1 - p_{vi1})}{(1 - p_{vi2})p_{vi1}} \right] = \eta + \tau - (\sigma'_i - \sigma_i) \quad (v = n + 1, \dots, N) \quad (16)$$

reveals that together with τ , the difference d_i of the item difficulties, $d_i = (\sigma'_i - \sigma_i)$, appears in both ratios as a further parameter. Thus, τ and d_i are confounded, so that neither the differences of the item difficulties nor the trend can be estimated separately unless the item parameters are known from an earlier investigation; if the parameters are known, σ_i and σ'_i can be considered constants in the model equations, and τ as well as η can be computed as for the standard LLRA.

If the LLRA is applied to each item I_1, \dots, I_k separately, for unknown σ_i and σ'_i , only the treatment effects η_i —now depending on the item—and the differences $\tau_i^* = \tau_i - d_i$ can be estimated. (Observe

that in this case the “true” trend, τ_i , also depends on the item.) But the interpretation of the latter is nearly impossible, for example with respect to the partitioning of the sample: If the difference τ_i^* is constant over subgroups of persons, it cannot be concluded that the trend τ_i and the difference d_i of the item difficulties are constant; if, on the other hand, the difference τ_i^* changes with the subgroups, it is not clear whether τ_i is constant while d_i changes or d_i is constant while τ_i changes. Similar problems arise when the constancy of the trend across the items is to be investigated for unknown item parameters.

If the LLRA is applied to all items I_1, \dots, I_k simultaneously, the corresponding hypothesis states that the confounded trends τ_i^* do not depend on the items, that is, $\tau_1 - d_1 = \dots = \tau_k - d_k$. In case this hypothesis can be maintained on the basis of appropriate statistical tests, τ_i and d_i may always sum to the same result; hence it is impossible to be certain that both of them do not change with the items. But it seems highly unrealistic to assume that such compensatory effects occur purely accidentally.

Caution must be exercised with respect to the trend parameters and their interpretation. Fortunately, this disadvantage will not cause great problems because the main goal of studies which intend to investigate change usually is not the quantification of the trend. Therefore, interest will focus on the treatment’s effects when applying the hybrid variant of the LLRA without having previous information about the item difficulties.

A similar problem of confounding occurs for the standard LLRA. But because this fact presents no consequences, it was ignored when presenting the simpler variant. In principle, fully parameterized, the specification equations of the standard LLRA can be written like those for the hybrid LLRA, except that $\sigma'_i = \sigma_i$ at t_2 in Equations 11 through 14. Thus, ξ_{vi} and σ_i are confounded, but can be eliminated together in the log odds ratios analogous to Equations 15 and 16, as well as when estimating the structural parameters of interest, η and τ , by the CML method. Therefore, it can be said that the concept of item difficulty does not initially appear in the hybrid LLRA, but is already implicit in the standard LLRA. However, in this latter model, item difficulty σ_i and person parameter ξ_{vi} are statistically indistinguishable because the same items are presented at both points in time to the same persons, who are characterized by time-invariant person parameters.

Because the person parameters ξ_{vi} ($v = 1, \dots, N; i = 1, \dots, k$), as incidental parameters, are eliminated both in the standard as well as in the hybrid LLRA during CML parameter estimation—which best is performed by specific computer programs, but not by adapting the programs available for parameter estimation in the usual linear logistic test model (LLTM; see, e.g., Fischer, 1983)—usually no information is obtained with respect to the performance level of the persons. Because for each person \times item combination only one binary response is available per point in time, separate and direct (unconditional maximum likelihood) estimates of the person parameters are also impossible.

However, rough estimates can be given to describe the performance of the persons, assuming homogeneity of the items and/or of the persons within well-defined subgroups, for example formed by treatment \times gender \times age $\times \dots$: For each item separately or summed over them, the frequencies of positive reactions ($+f_g$) and negative reactions ($-f_g$) at t_1 are counted per subgroup, S_g , and then transformed according to the logistic function into the subgroup-specific performance parameters $\bar{\xi}_g$, that is, $\bar{\xi}_g = \log(+f_g/-f_g)$. They give the approximate average of the performance at t_1 of those persons contained in subgroup S_g ; hence not only can the changes (in terms of η and τ) be quantified by the LLRA, but also—at least approximately—the starting positions of the persons; see below.

Empirical Example

Method

As a first application of the hybrid variant of the LLRA, an empirical investigation concerning the effects of training text comprehension is presented. In the pretest, students about 13 years of age in four

school classes were given a short story to read. After reading the story, the students completed the following four tasks: (1) labeling the steps of the story, (2) labeling the key passages of the story, (3) writing a summary, and (4) identifying the message of the story. Then, for the following three lessons, each of the four classes was assigned at random to a particular experimental condition:

Class 1 (control group, $n_1 = 30$) was not trained in text condensation and was not confronted with similar stories.

Class 2 (experimental group 1, $n_2 = 26$) was trained by means of two similar stories on grammatical rules but not on text condensation.

Class 3 (experimental group 2, $n_3 = 22$) was trained in the analysis of texts following the principles of Thorndyke (1977) and Kintsch and van Dijk (1978). Whereas Thorndyke structured a story into frame, theme, complication, and solution, Kintsch and van Dijk gave the following rules for text condensation: generalization (finding the generic term from several statements), selection (of the typical statement out of a number of them), omitting (of statements which do not contribute to the coherence of the text), and construction (formulating a new statement which can substitute for a number of statements in the text). First, the students were confronted with these rules, and then they trained in them by applying them to two stories.

Class 4 (experimental group 3, $n_4 = 24$) was trained like Class 3, and, in addition, was asked to practice the describing of pictures: Classmates described the pictures they actually saw to two of their colleagues, who then had the task of drawing the pictures vicariously, in the best approximation possible of the descriptions. Then the copies were compared to the originals so that the consequences of missing and inexact transfer of information became apparent.

In the posttest, all students had to accomplish the same four tasks as in the pretest, but on another story of about equal length and structure which was rated by experts to be more difficult compared to the story of the pretest.

After scoring the students' answers dichotomously (correct vs. incorrect answers), a number of hypotheses were formulated concerning the effect of the three experimental conditions, and the trend which was confounded with the difference of the difficulties of each item with respect to the two different stories used at both points in time.

Then, these hypotheses were contrasted to the saturated one, which in the present study assumed that the changes were dependent on the treatment conditions and on the gender of the students (this was the sole control variable used to test the generalizability of the effects over the persons), by means of the likelihood ratio statistic

$$\lambda = -2 \log (L/L^*) = 2(\log L^* - \log L) \quad , \quad (17)$$

which is asymptotically χ^2 distributed with degrees of freedom $= h^* - h$. L^* denotes the likelihood under the saturated hypothesis H^* assuming h^* parameters, and L denotes the likelihood under a somewhat restricted hypothesis H with h parameters, $h < h^*$.

Because experimental conditions were used in case 4, the assumed interaction with the control variable "gender" leads to a saturated hypothesis with eight parameters whose relation to the log odds ratios (Equations 15 and 16) of the diverse subgroups of persons is given below, where the symbol τ^* is used to indicate the confounding of the trend with the differences of the item parameters.

Control Group	females: $\ell_{1f} = \tau_f^*$
	males: $\ell_{1m} = \tau_m^*$
Experimental Group 1	females: $\ell_{2f} = \tau_f^* + \eta_{1f}$
	males: $\ell_{2m} = \tau_m^* + \eta_{1m}$

$$\begin{array}{ll}
 \text{Experimental Group 2 females: } \ell_{3f} = \tau_f^* & + \eta_{2f} \\
 \text{males: } \ell_{3m} = \tau_m^* & + \eta_{2m} \\
 \\
 \text{Experimental Group 3 females: } \ell_{4f} = \tau_f^* & + \eta_{3f} \\
 \text{males: } \ell_{4m} = \tau_m^* & + \eta_{3m}
 \end{array} \tag{18}$$

Each of the restricted hypotheses (see Table 1) is derived from the saturated one by equating or deleting certain parameters. For example, the hypothesis H_2 of no trend (i.e., $\tau_f^* = \tau_m^* = 0$) but gender-specific effects of the treatment conditions results from deleting the trend parameters from Equation 18; the hypothesis H_6 of no trend and identical effects of the treatment conditions for both genders results from additionally equating $\eta_{1f} = \eta_{1m} = \eta_1$, $\eta_{2f} = \eta_{2m} = \eta_2$, and $\eta_{3f} = \eta_{3m} = \eta_3$; and the hypothesis H_{12} of no trend and gender-independent identical effect of all treatment conditions results from equating $\eta_1 = \eta_2 = \eta_3 = \eta$. Thus, under H_2 six parameters must be estimated, under H_6 three parameters, and under H_{12} one parameter.

Results

As the likelihood ratio tests (LRTs) in Table 1 show, the most parsimonious hypothesis with a nonsignificant test statistic is H_2 ($\lambda = .90$, $df = 2$, $\chi_{95}^2 = 5.99$) while hypotheses such as H_6 and H_{12} are rejected. Thus it can be concluded that, depending on the gender of the students, each of the three

Table 1
Hierarchy of Hypotheses Concerning the Changes
Attributable to the Treatments and the Trend,
Assumptions for η and τ , Number of Parameters Estimated,
and Results of the Likelihood Ratio Tests (LRT)

Hypothesis	η	τ	Param- eters	log L	LRT		χ_{95}^2
					H_1 vs. H^*	df	
H^*	G×T	G	8	-103.61	--	-	--
H_1	G×T	+	7	-103.85	.48	1	3.84
H_2	G×T	-	6	-104.06	.90	2	5.99
H_3	T	G	5	-108.70	10.18	3	7.82
H_4	T	+	4	-110.00	12.78	4	9.49
H_5	G	G	4	-111.07	14.92	4	
H_6	T	-	3	-110.21	13.20	5	11.07
H_7	+	G	3	-111.11	15.00	5	
H_8	G	+	3	-111.31	15.40	5	
H_9	+	+	2	-111.50	15.78	6	12.59
H_{10}	G	-	2	-111.52	15.82	6	
H_{11}	-	G	2	-111.69	16.16	6	
H_{12}	+	-	1	-111.71	16.20	7	14.07
H_{13}	-	+	1	-112.09	16.96	7	

Note. G×T indicates that the effect parameters were assumed to be dependent upon the combination of gender and treatment (interaction). G and T indicate that the parameters were assumed to be dependent upon gender or treatment. "+" indicates that a single parameter τ (not dependent upon gender) or η (not dependent upon the treatment and gender) was assumed. "-" indicates that no parameter was assumed.

experimental conditions had its own effect on text comprehension which was operationalized here by four items; the trend, which was confounded with the difference of the difficulty of both texts used at t_1 and t_2 , can be neglected. These findings are based on the simultaneous analysis of all four items assuming the constancy of the parameters over them. Therefore, a further statistical test based on the likelihood ratio statistic (Equation 17) was needed to justify this assumption.

Denote by L the maximized likelihood of the data comprising all items I_1, \dots, I_k under a certain hypothesis H underlying the changes and providing for h parameters $\eta_1, \dots, \eta_{h-1}, \tau$. Let L_i be the likelihood maximized separately for item I_i under the same hypothesis H . Thus, L is a function of $\eta_1, \dots, \eta_{h-1}, \tau$, and L_i is a function of the item-specific parameters $\eta_1^{(i)}, \dots, \eta_{h-1}^{(i)}, \tau^{(i)}$.

Combining the L_i s to L^* ,

$$L^* = \prod_{i=1}^k L_i, \quad (19)$$

gives the likelihood of all items, but now depending on $\eta_1^{(1)}, \dots, \eta_{h-1}^{(1)}, \tau^{(1)}, \dots, \eta_1^{(k)}, \dots, \eta_{h-1}^{(k)}, \tau^{(k)}$. To test the hypothesis that the effect parameters are constant over the items, L^* and L can be contrasted with each other by the likelihood ratio statistic Equation 17. The degrees of freedom are $df = h(k - 1)$, because L depends on h parameters and L^* is a function of kh parameters.

For H_2 , this statistical test was not significant ($\lambda^* = 24.46$, $df = 18$, $\chi_{95}^2 = 28.87$), indicating the generalizability of the gender-specific treatment effects over the four items. The parameter estimates given in Tables 2 and 3 reveal relatively large discrepancies between boys and girls as well as between the three treatment conditions. Positive effects of about equal size were produced by treatment condition 1 (no training, but confrontation with similar stories) for boys and girls, while treatment condition 2 (training on texts) led to marked positive effects for boys but small negative effects for girls; the opposite held for treatment condition 3 (training on texts and on pictures), causing about equally large changes in the intended direction for girls and in the unwelcome direction for boys.

Considering for descriptive purposes the mean effects of the treatments over boys and girls, that is, ignoring the interaction of treatment \times gender—which, strictly speaking, is inadmissible—the effect of the first treatment was positive while the effects of the two other treatments were about 0; similarly, considering the mean of the treatment effects for both genders (i.e., averaging the effects over the treatments separately for boys and girls), the boys did not profit from the training whereas the girls did (see the marginals in Table 2). This fact can only be observed if no trend is assumed, as is the case under H_2 . For the saturated hypothesis H^* (see Table 3), for the boys the inefficiency of the treatments found under H_2 was caused by their negative trend, which of course was not significant; otherwise the test

Table 2
Parameter Estimates for the Treatment Effects and the
Trend Under Hypotheses H_2 , H_6 , H_{10} , and H_{12}

H_2	Boys	Girls	Means Over Genders (H_6)
τ	--	--	--
η_1	.547	.693	.595
η_2	1.386	-.405	-.172
η_3	-.916	1.041	.091
Means Over Treatments (H_{10})	.071	.296	.194 ^a

^aMean over treatments and gender (H_{12}).

Table 3
 Parameter Estimates for the Treatment Effects and the
 Trend Under Hypotheses H^* , H_4 , H_5 , and H_9

H^*	Boys	Girls	Means Over Genders (H_4)
τ	-.452	.000	-.211
η_1	.999	.693	.806
η_2	1.838	-.405	.039
η_3	-.464	1.041	.302
Means Over Treatments (H_5)	.523	.296	.405 ^a

^aMean over treatments and gender (H_9).

statistic for H_2 could not remain insignificant. But it can now be seen that over the genders as well as over the treatments all mean effects were positive. Thus, on the average, the training was suitable to improve text comprehension.

This somewhat surprising result may be clarified by the following comments:

1. Each experimental condition corresponded to a school class, so that neither a general interaction nor a gender-specific interaction of teacher and treatment (i.e., class) can be precluded (the teachers in Classes 1 and 4 were female, and those in Classes 2 and 3 were male).
2. The sample sizes were relatively small ($n_1 = 30$, $n_2 = 26$, $n_3 = 22$, $n_4 = 24$); thus the changes in the dichotomous data (from “+” to “-”, and reverse) were not numerous and the accuracy of the parameter estimates decreased.
3. The dichotomization of the answers of the students may have been too rough; perhaps, splitting of each of the items into more components would have been more informative, leading to more data per person.

In addition to the effects of the treatments, the performance levels of the different subgroups may be of further interest. Because of the small sample sizes, they were computed according to the description given above on the basis of the collapsed data (i.e., for all items together); see Table 4. Noticeable differences in the performance level can be found within the same class for different genders (except Class 1) but also between the classes (Classes 1 and 2 vs. Classes 3 and 4); on the average, boys and girls were rather similar with respect to their starting positions in the present experiment, whereby the

Table 4
 Frequencies of Positive (+f) and Negative (-f)
 Reactions for the Treatment \times Gender Groups,
 and Their Performance Levels at t_1

Group	Frequency				Performance Level at t_1		
	Boys		Girls		Boys	Girls	Total
	+f	-f	+f	-f			
Class 1	15	45	17	43	-1.099	-.928	-1.012
Class 2	12	52	13	27	-1.466	-.731	-1.151
Class 3	6	18	27	37	-1.099	-.351	-.511
Class 4	21	23	16	36	-.091	-.811	-.467
Total	54	138	73	143	-.938	-.672	-.794

total mean of performance level, $-.794$, and its corresponding probability for positive reactions, $.311$, show that the four items, on the average, were rather difficult to solve for the students at t_1 .

Conclusions

The present article reports the first application of the hybrid variant of the LLRA probabilistic model to the measurement of change in qualitative (dichotomous) data. In light of previous multiple applications of the standard version of the LLRA (see Fischer & Formann, 1982), the following conclusions can be drawn.

1. In cases where the standard LLRA possibly fails (the presentation of the same items twice seems to be problematic; the trend and/or the treatment effects are presumably too massive), the hybrid LLRA can be a useful tool to quantify the effects of certain experimental conditions. Like the more simple standard variant, it allows the formulation of a variety of hypotheses with respect to the changes and their (in)dependence on the treatments, the persons, and the items. Powerful likelihood ratio tests permit the statistical evaluation of these hypotheses to determine whether the regularities are inherent in the changes and to investigate the generalizability of the treatment effects across the items and the persons.
2. Whenever it may be practical, the investigation should be performed in two independent steps, the first to estimate the parameters of those items which will be used at both points in time, and the second to assess the changes. Otherwise, the item parameters are confounded with the trend so that only the effect parameters of the treatments can be interpreted. This was the case in the above empirical example.
3. In regard to the study on training of text comprehension, in addition to this problem some other weaknesses in the experimental design could be identified to which the somewhat unexpected results (partially negative effects of the training conditions) could be attributed. They are the possible interaction of teacher and treatment, the small sample sizes, and the modest number of items at both points in time.

Appendix

The estimates of the effect parameters of the LLRA diverge when changes in the manifest data are too great. The first step in demonstrating this is the formulation of the conditional maximum likelihood equations.

The CML Equations of the LLRA

Let a_{vi1} and a_{vi2} denote the dichotomously scored answers of person P_v for item I_i at t_1 and t_2 , respectively, where 1 indicates a positive reaction and 0 a negative one. Then L_{vi}^* , the likelihood of the answers of person P_v for item I_i at both points in time, equals

$$L_{vi}^* = p(a_{vi1}, a_{vi2}) = \frac{\exp(a_{vi1}\xi_{vi})}{1 + \exp(\xi_{vi})} \cdot \frac{\exp[a_{vi2}(\xi_{vi} + \delta_v)]}{1 + \exp(\xi_{vi} + \delta_v)} \quad (A1)$$

where

$$\delta_v = \sum_{j=1}^{m+1} q_{vj}\eta_j, \quad (A2)$$

that is, the trend τ is the $(m + 1)$ th effect parameter η_{m+1} whose treatment amount $q_{v,m+1}$ is equal to 1 for $v = 1, \dots, N$. (For a schematic representation of the data and the treatment amounts, see Table 5.)

Next, consider the conditional likelihood of the same answers given their sum $a_{vi0} = a_{vi1} + a_{vi2}$. Four cases must now be distinguished, namely $p(0,0|a_{vi0} = 0)$, $p(0,1|a_{vi0} = 1)$, $p(1,0|a_{vi0} = 1)$, and $p(1,1|a_{vi0} = 2)$.

Table 5
 Data and Treatment Doses

Persons	Points in Time		Treatments 1...j...m	Trend m+1
	t_1	t_2		
	Items 1...i...k	Items 1...i...k		
1	.	.	.	1
.
.
.
v . . .	a_{vi1} . . .	a_{vi2} . . .	q_{vj} . . .	1
.
.
N	.	.	.	1

= 2). Because both $p(0,0|a_{vi0} = 0)$ and $p(1,1|a_{vi0} = 2)$ equal 1—in both cases, for the sum $(a_{vi1} + a_{vi2})$, exactly one combination of 0s and 1s exists—and because substituting Equation A1 for $p(0,1|a_{vi0} = 1)$ and $p(1,0|a_{vi0} = 1)$ leads to

$$p(0,1|a_{vi0} = 1) = \frac{\exp(\delta_v)}{1 + \exp(\delta_v)} \tag{A3}$$

and

$$p(1,0|a_{vi0} = 1) = \frac{1}{1 + \exp(\delta_v)} \tag{A4}$$

which can be combined into

$$p(a_{vi1}, a_{vi2}|a_{vi0} = 1) = \frac{\exp(a_{vi2}\delta_v)}{1 + \exp(\delta_v)} \tag{A5}$$

the conditional likelihood, L_{vi} , for all four cases can be written as

$$L_{vi} = p(a_{vi1}, a_{vi2}|a_{vi0} = 1) = \left[\frac{\exp(a_{vi2}\delta_v)}{1 + \exp(\delta_v)} \right]^{(a_{vi1} - a_{vi2})^2} \tag{A6}$$

The conditional likelihood, L , for all persons and items equals

$$L = p(\mathbf{A}_1, \mathbf{A}_2|\mathbf{A}_1 + \mathbf{A}_2) = \prod_v \prod_i L_{vi} \tag{A7}$$

where $\mathbf{A}_1 = \|a_{vi1}\|$ and $\mathbf{A}_2 = \|a_{vi2}\|$ are the item-score matrices at t_1 and t_2 , respectively.

Taking logarithms, differentiating, and setting equal to 0 gives the CML estimation equations of the LLRA parameters (see Fischer, 1983, Equation 48):

$$\sum_v \sum_i (a_{vi1} - a_{vi2})^2 q_{v\ell} \left[a_{vi2} - \frac{\exp\left(\sum_{j=1}^{m+1} q_{vj}\eta_j\right)}{1 + \exp\left(\sum_{j=1}^{m+1} q_{vj}\eta_j\right)} \right] = 0 \quad (\ell = 1, \dots, m+1) \tag{A8}$$

Parameter Estimates for Massive Changes in the Manifest Data

According to Equation A8 for CML parameter estimation, for a given item only that part of the data is used for which changes were observable [i.e., for which $(a_{vi1} - a_{vi2})^2 = 1$]. After exclusion of the uninformative data with $a_{vi1} = a_{vi2}$, Equation A8 can therefore be rewritten as

$$\sum_v q_{v\ell} a_{v02} = \sum_v k_v q_{v\ell} \frac{\exp\left(\sum_{j=1}^{m+1} q_{vj} \eta_j\right)}{1 + \exp\left(\sum_{j=1}^{m+1} q_{vj} \eta_j\right)} \quad (\ell = 1, \dots, m+1) \quad , \quad (\text{A9})$$

where $a_{v02} = \sum_i a_{vi2} (a_{vi1} - a_{vi2})^2$ is the score of person P_v at t_2 for the remaining informative items (i.e., those for which the answers changed from t_1 to t_2), and k_v is their number.

For simplicity, consider again the case of only two groups of persons, the first being the control group with persons P_1, \dots, P_n , and the second being the experimental group with persons P_{n+1}, \dots, P_N . This gives the following two equations which must be solved in order to estimate η and τ :

$$\sum_{v=n+1}^N a_{v02} = \sum_{v=n+1}^N k_v \frac{\exp(\eta + \tau)}{1 + \exp(\eta + \tau)} \equiv \sum_{v=n+1}^N k_v \pi_1 \quad \text{for } \eta \quad , \quad (\text{A10})$$

$$\sum_{v=1}^n a_{v02} = \sum_{v=1}^n k_v \frac{\exp(\tau)}{1 + \exp(\tau)} + \sum_{v=n+1}^N k_v \frac{\exp(\eta + \tau)}{1 + \exp(\eta + \tau)} \equiv \sum_{v=1}^n k_v \pi_2 + \sum_{v=n+1}^N k_v \pi_1 \quad \text{for } \tau \quad . \quad (\text{A11})$$

If the manifest effect of the treatment is extremely massive in the sense that all persons of the experimental group change from negative (0) answers at t_1 to positive (1) answers at t_2 , then the score at t_2 , a_{v02} , equals the number of informative items, k_v , for $v = n + 1, \dots, N$, and the right side of Equation A10 can only be equal to the left side if $\pi_1 = 1$. If, on the contrary, all persons of the experimental group change from positive answers at t_1 to negative answers at t_2 , then $a_{v02} = 0$ for $v = n + 1, \dots, N$, and π_1 must be equal to 0.

Thus, in case of changes solely in one direction in the experimental group—while changes occur in both directions in the control group—the parameter estimate for η tends to $\pm\infty$ during the iterative solution of the CML estimation equations. By analogy, the same holds with respect to τ , assuming massive effects in the control group, and also more generally for $q_{vj} \notin \{0, 1\}$.

References

- Breslow, N. (1976). Regression analysis of the log odds ratio: A method for retrospective studies. *Biometrics*, 32, 409–416.
- Duncan, O. D. (1985a). New light on the 16-fold table. *American Journal of Sociology*, 91, 88–128.
- Duncan, O. D. (1985b). Some models of response uncertainty for panel analysis. *Social Science Research*, 14, 126–141.
- Fischer, G. H. (1972). A measurement model for the effect of mass-media. *Acta Psychologica*, 36, 207–220.
- Fischer, G. H. (1976). 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.
- Fischer, G. H. (1977). Linear logistic latent trait models: Theory and application. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning*. Bern: Huber.
- Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3–26.
- Fischer, G. H., & Formann, A. K. (1982). Veränderungsmessung mittels linear-logistischer Modelle. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 75–99.
- Harris, C. W. (Ed.). (1963). *Problems in measuring change*. Madison: University of Wisconsin Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

- Koch, G. G., Landis, R. J., Freeman, J. L., Freeman, D. H., Jr., & Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, *33*, 133–138.
- Marascuilo, L. A., & Serlin, R. D. (1979). Tests and contrasts for comparing change parameters for a multiple sample McNemar data model. *British Journal of Mathematical and Statistical Psychology*, *32*, 105–112.
- Plewis, I. (1981). A comparison of the approaches to the analysis of longitudinal categorical data. *British Journal of Mathematical and Statistical Psychology*, *34*, 118–123.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research. [Expanded edition, University of Chicago Press, 1980.]
- Thorndyke, P. W. (1977). Cognitive structure in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, 77–110.

Author's Address

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